

Modeling and Estimation of Selection Interests through Gaze Behavior

Kei Shimonishi

Abstract

Multiple criteria decision making (MCDM) is a fundamental part of our daily lives. To support solving the problem with conflicting multiple criteria, several analysis methods for MCDM are hitherto proposed. Here, the choice of evaluation criteria is a key to successful decision-making support, including interactive assistance systems. With appropriate evaluation criteria, decision-support systems can help users find an importance weight on each of the criteria and organize their selection interests. This process of preference structuring is helpful for users who want to select a target from alternatives, particularly when they have uncertain preferences and have not yet detailed their needs enough to search targets using appropriate keywords. Here, the questions are how to prepare evaluation criteria beforehand and how to estimate user preferences by using these evaluation criteria.

This thesis introduces *aspects* that represent “why the users look at items,” which provide possible viewpoints of items and proposes a framework for obtaining aspects from users’ gaze behavior while browsing multi-attribute content.

In the proposed framework, we introduce an aspect-oriented probabilistic generative model of gaze behavior that is based on topic models. The proposed gaze behavior model learns aspects in a data-driven fashion as a form of the degree of association with each attribute value. Also, this model estimates users’ selection interests by using learned aspects. By estimating users’ selection interests through constructing state space of selection interests from a large amount of users’ gaze behavior, this model can be applied to proactive item recommendation.

The following two characteristics in this situation make the design of the gaze behavior model difficult: 1) users do not always compare items in the content, and users do not always focus on all attributes of items; and 2) users’ gaze behavior is affected by the spatial layout of the content. This thesis proposes two kinds of models to overcome these difficulties.

For the first difficulty, this thesis introduces a multiscale detection method of users' comparison behavior named the *Multiscale Exact Test (MSET)*. The idea of MSET is that users' significant gaze behavior can be detected using the deviations in the distributions of gaze observations by modeling users' neutral gaze behavior. Users' comparison behavior can be detected together with its scale by comparing the p -value calculated from observed gaze behavior to the significance level. By regarding the attribute-of-focus detected by the MSET as the observations of generative gaze behavior model, the proposed generative model learns aspects taking into account users' comparison behavior. We evaluated the validity of this two-step framework of aspect learning using the correlations between learned aspects and task-related attribute values.

For the second difficulty, this thesis specifically focuses on the spatial effect of users' gaze behavior called center bias. The basic idea of this approach is to unify both external factors (i.e., center bias) and internal factors (i.e., preference) in a single generative model. The validity of this proposed framework is evaluated by the prediction accuracy of the region of interests or item of interests through selection interests.

Once aspects are learned, the learned aspects can be regarded as evaluation criteria of alternatives of MCDM and can be used for an interactive system that assists users' MCDM such that the system assists users to recognize the problem structure by directly mentioning aspects. This thesis finally introduces aspect-based interactive assistance using the hierarchical structure of decision problem. By estimating user preferences from gaze behavior and mentioning aspects of alternatives, the system helps users organize their selection interests and decide an alternative to choose. Through interaction during decision making, we show the validity of the effectiveness of using aspects.

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor Associate Professor Hiroaki Kawashima for providing me valuable advice and good opportunities to research on this topic.

I really appreciate Professor Takatsune Kumada and Professor Tatsuya Kawahara for taking time to review my thesis and giving me helpful comments about my study.

I would also like to express my sincere appreciation to Professor Takashi Matsuyama. When I was undergraduate and at master's courses, he was my supervisor, and he invited me to this interesting research and provided insightful advice.

I am also thankful for Dr. Shohei Nobuhara and Dr. Ryo Yonetani for giving me a lot of advice and constant supports and encouragements. Special thanks go to our secretaries, Ms. Aya Inoue, Ms. Asako Yoshimura, and Ms. Chieko Funato for their assistance of office job, and to all the former and current members of Matsuyama Laboratory, especially but not limited to Dr. Erina Ishikawa Schaffer and Mr. Ryo Kawahara.

I have been blessed with many like-minded colleagues, Mr. Yoshiaki Bando, Mr. Koji Inoue, Mr. Tomohiro Sakaguchi, and Mr. Motoharu Sonogashira in the same department, Graduate School of Informatics Department of Intelligence Science and Technology, and Mr. Norihito Fukui, Mr. Hirotaka Kato, Mr. Jun Miyanohara, Mr. Takuya Ohmura, Ms. Yoko Okajima in the same committee of Kyoto University.

Finally, I wish to express my gratitude to my family.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Users' Choice Behavior	2
1.2.1	Utilities	2
1.2.2	Multiple Criteria Decision Analysis	2
1.2.3	Analytic Hierarchy Process	5
1.2.4	User Preferences in Choice Behavior	6
1.2.5	Phase of Choice Behavior	6
1.3	Selection Assistance	7
1.3.1	Component in Selection Assistance	7
1.3.2	Information Presentation to Users	8
1.3.3	Estimation of Users' Internal States	8
1.3.4	Information Acquisition to Assistants	9
1.3.5	Information Conveyance to Assistants	9
1.3.6	Interaction During Assistance	9
1.3.7	Interactive Decision-Support System	10
1.4	Gaze Behavior in Catalog Content Browsing	11
1.4.1	Features of Eye Movements	11
1.4.2	The Factors of Eliciting Users' Gaze Behavior	12
1.4.3	Users' Gaze Behavior and Decision Making	13
1.5	Problem Statement	13
1.6	Contributions	15
1.6.1	Representation of users' internal states	15
1.6.2	Users' comparison behavior	15
1.6.3	Gaze Behavior Modeling	16
1.7	Structure of the Thesis	16

2	Framework of Aspect-Oriented Gaze Behavior Model	19
2.1	Introduction	19
2.2	Obtaining State Space of Users' Internal States from Users' Behavior	20
2.2.1	Learning State Space by Topic Models	21
2.3	Learning Evaluation Criteria from Gaze Behavior	22
2.4	Component of the Gaze Behavior Model	22
2.4.1	Aspects	22
2.4.2	Representation of Users' Internal States	23
2.5	Formulations of the Content-Browsing Situation	24
2.5.1	Model of Aspects	24
2.5.2	Model of Selection Interests	25
2.6	Generative Process of Gaze Behavior	25
2.6.1	Observation model	27
2.7	Estimation of Model Parameters	28
2.7.1	Learning Aspects (Figure 2.4 (a))	29
2.7.2	Estimating Selection Interests (Figure 2.4 (b))	29
2.7.3	Predicting of Observation (Figure 2.4 (c))	30
2.8	Extension of the Propose Model	30
3	Multiscale Exact Test: Comparison Behavior Detection using Statistical Testing	31
3.1	Introduction	31
3.2	The Concept of the Multiscale Exact Test	32
3.3	Multiscale Exact Test	35
3.3.1	In Case of Categorical Attributes	35
3.3.2	In Case of Ordinal Attributes	37
3.4	Evaluation	40
3.4.1	Toy Data	40
3.4.2	Results on Toy Data	41
3.4.3	Actual Gaze	43
3.4.4	Results on Gaze Data	46
3.5	Discussion	48
4	Gaze Behavior Model with Attribute-of-Focus for Aspect Learning	51
4.1	Introduction	51
4.2	Approach for Learning Evaluation Criteria	52
4.3	Formulations of the Content-Browsing Situation	54

4.4	Generative Process of Attribute-of-Focus	55
4.5	Estimation of Model Parameters	56
4.6	Evaluation	56
4.6.1	Toy Data	58
4.6.2	Actual Gaze Data	59
4.7	Discussion	63
5	Gaze Behavior Model with the Spatial Effect for Item Recommendation	67
5.1	Introduction	67
5.2	Generative Process of Gaze Behavior	68
5.3	The Effect of Layout in the Gaze Behavior Model	69
5.4	Estimation of Model Parameters	71
5.5	Model of Effect of Layout	72
5.5.1	Mixture Model	72
5.5.2	Prior-Based Model	73
5.5.3	Difference of Two Models	74
5.6	Evaluation	74
5.6.1	Experimental Setting	75
5.6.2	Procedure	77
5.6.3	Evaluation Scheme	78
5.6.4	Results	81
5.7	Discussion	84
6	GazeAHP: Tracing Selection Interests and Supporting Users to Recognize the Structure of Problems	87
6.1	Introduction	87
6.2	The framework of GazeAHP	88
6.3	Interaction design	89
6.4	Tracing Temporal Changes of Selection Interests	90
6.5	Evaluation	91
6.5.1	Experimental Design	92
6.5.2	Procedure	93
6.5.3	Analysis and Discussion	93
7	Conclusion	101
7.1	Summary	101
7.2	Future Work	102

7.2.1	Limitations and Extensions	102
7.2.2	Prospects to Interactive Decision Assistance	104

List of Figures

1.1	Categorization of the domain of choice behavior	4
1.2	An example of structure of the AHP	5
1.3	Component in selection assistance	8
1.4	Structure of the thesis	16
2.1	Graphical model of the pLSA	21
2.2	An example of aspects	23
2.3	Hierarchical structure for decision making. Users' selection interests are represented as each evaluation criterion's weight of importance, and evaluation criteria are characterized by how they are related to each attribute value. Attribute values and alternatives are connected if alternatives have the specific attribute value.	24
2.4	Generative process of gaze behavior from users' selection interests	25
2.5	Graphical model of the pLSA and the proposed model	27
3.1	An example of detected attribute-of-focus. Each row shows an attribute value. Each column shows one attribute-value vector of browsed target, and white color denotes an item has that attribute value. (a) shows an original attribute-value sequence, (b) shows a sequence of attribute-of-focus detected by MSET.	33
3.2	Proposed flow of the MSET to detect attribute-of-focus for categorical attributes. Different color in attribute-of-focus denotes different attribute value.	36
3.3	Proposed flow of the MSET to detect attribute-of-focus for ordinal attributes. Different color in attribute-of-focus denotes different attribute value.	38

3.4	An example of the results of the MSET on toy data: (a) p -values and (b) identified distinctive browsing period and detected attribute-of-focus. Each figure of (a) shows when and how long distinctive browsing occurs. Since attribute type Y and Z are ordinal attributes, three figures of p -values are shown for three thresholds. Each figure of (b) corresponds to each attribute type, and different colors indicate different attribute values. White and gray periods show neutral browsing phase and focused browsing phases, respectively.	42
3.5	Similarities between detected attribute of focus and task related attribute values	43
3.6	Experimental environment	44
3.7	Points used for the calibration of the eye tracker. This figure express only where these points are located, and size is exaggerated. .	45
3.8	(a) Histogram of session duration length. The mean of duration length was 33.3 seconds and the median was 25.6 seconds.(b) Histogram of the number of gaze target transitions in each session. The mean number of times was 37.8, and the median was 33. . . .	47
3.9	(a) Histogram of the ratio of focused browsing phase in each session detected by the MSET. The mean of the ratio was 64.9%, and the median was 69.4%. Note that these durations were not actual sampling times but were based on the number of times targets were looked at in a session. (b) Histogram of the number of attribute values of focus in attribute-of-focus detected by the MSET for all session. The mean number of focused value was 1.59, and the median was 2.	47
3.10	An example of the results of the MSET on gaze data: (a) p -values and (b) identified distinctive browsing period and detected attribute-of-focus. Each figure of (a) shows when and how long distinctive browsing occurred. Since all attributes are ordinal, two figures of p -values for two thresholds. Each figure of (b) corresponds to each attribute type.	49
3.11	Similarities between detected attribute of focus and task related attribute values from actual data	50

LIST OF FIGURES

4.1	Proposed flow of two-step approach to obtain evaluation criteria from users' gaze behavior. The first step detects users' attribute-of-focus, and the second step learns evaluation criteria from detected attribute-of-foci.	52
4.2	Graphical models of pLSA and proposed model	53
4.3	A probabilistic generative process of attribute-of-focus from users' selection interests	53
4.4	Given tasks and learned aspects with and without the MSET on toy data. The size and color of each dot depict the value of the multinomial parameter $p_{q,k}$, which corresponds to the degree the aspect C_q is related to attribute values V_k	57
4.5	Similarities between given evaluation criteria and learned aspects with and without the MSET. Each bar of the (a) shows the similarity between learned aspect and the evaluation criterion given as the task. (b) shows the mean of similarities of all tasks.	57
4.6	An example of task-related attribute values (given evaluation criteria) and learned aspects with and without the MSET on the gaze data. The number of aspects was set to be three, the same as that of tasks. The size and color of each dot depict the value of the multinomial parameter $p_{q,k}$, which represents the degree the aspect C_q is related to attribute values V_k	60
4.7	Similarities between task-related attribute values and aspects with and without the MSET. Each bar of (a) shows the similarity between learned aspect and each evaluation criterion given as the task on actual gaze data. (b) shows the mean of similarities of all tasks.	61
4.8	Learned aspects by changing the number of aspects	64
5.1	Generative process of gaze behavior from users' selection interests	68
5.2	Normalized frequency distribution of region-of-gaze on the display obtained from the experiment in Section 5.6	70
5.3	Graphical model of the pLSA and the proposed model	70

5.4	Example of displayed content. Every image is diverted from a web page, Mizkan Co., Ltd. (http://www.mizkan.co.jp/index.html), and a book (Sally, photo by Takeshi Kubota (2013). I scientificized “delicious” and made it a recipe (in Japanese). Sunmark.) by getting permissions for using in our experiments.	76
5.5	The procedure of the experiment	77
5.6	Flow chart of the comparison method	80
5.7	Prediction accuracy of region-of-gaze. 50[%] of each session was used to estimate selection interest.	81
5.8	Prediction accuracy of region-of-gaze with respect to different mixture ratio β (the number of aspects: 6, used attributes: SA) and different ratio in each of test sessions used for estimating the participant’s selection interests.	82
5.9	Prediction accuracy of item-of-interest. 50[%] of each session was used to estimate selection interest.	83
5.10	Prediction accuracy of item-of-interest with respect to different mixture ratio β (the number of aspects: 6, used attributes: SA) and different ratio in each of test sessions used for estimating the participant’s selection interests.	83
5.11	Examples of learned aspects ($\beta = 0.5$, used attributes: SA). Left: the number of aspects: 3, right: the number of aspects: 12. Attribute values with high score have strong relation with that aspect.	84
5.12	Examples of estimated selection interests $\theta(s)$ in each session ($\beta = 0.5$, used attributes: SA, displayed image set: 1). Sessions with high score have strong relation with that aspect shown in Figure 5.11 (right).	85
6.1	Interaction flow. Orange squares show the system’s action and blue diamond shapes shows the participants’ reaction.	90
6.2	Experiment environment	91
6.3	The number of times interaction occurs and the total duration of decision making	94
6.4	The ratio that the system’s probing action is correct	94
6.5	The number of times interaction occurs and the total duration of decision making only when all the estimated selection interests were collect during one choice behavior	95

LIST OF FIGURES

6.6	(a) An example of transition of estimated α_t and (b) the timing of system's action and participant's reaction, where + is yes, and – is no.	96
6.7	The histogram of transition of mean of α_q after system's action. (a) is a histogram of AB strategy and (b) is that of CB strategy.	96
6.8	The histogram of region-of-gaze in the all sessions. (a) The histogram before sorting action (i.e., first displayed content) of all session and (b) the histogram after sorting action in AB strategy, and (c) the histogram after sorting in CB strategy.	97
6.9	Subjective evaluation for the decision behavior	97
6.10	Subjective evaluation for interactions with the system	98
6.11	Subjective evaluation for load for the interaction	98

List of Tables

1.1	Categorization of interactive assistance system	10
3.1	Given tasks and attribute values expected to be related to the tasks	45
4.1	Results of evaluation criterion estimation on toy data. Number in each cell shows the number of sessions that were estimated as engaged task q ($q = 1, \dots, 4$). “No comparison” means the MSET could not detect a bias of gaze behavior in the session.	58
4.2	Results of evaluation criterion estimation on gaze data with the MSET. Number in each cell shows the number of sessions that were estimated as engaged task q ($q = 1, 2, 3$). “No comparison” means the MSET could not detect a bias of gaze behavior in the session. .	62
4.3	Results of evaluation criterion estimation on gaze data without the MSET. Number in each cell shows the number of sessions that were estimated as engaged task q ($q = 1, 2, 3$).	62
5.1	Image set used in each task	76

Chapter 1

Introduction

1.1 Motivation

Candidate selection is a fundamental part of our daily lives. For example, we routinely choose a restaurant to eat dinner or choose a camera to use. In such cases, people often evaluate candidates from several criteria, such as the average costs and distance from the current location.

However, people sometimes do not have enough knowledge about a content domain or have only fuzzy understanding on their preferences. Therefore, preparing appropriate evaluation criteria tends to be difficult. Also, when people have not yet detailed their needs enough, they experience difficulties in giving appropriate assessment scores to each evaluation criterion [Chowdhury et al., 2011, Mahmood and Ricci, 2009]. These difficulties cause a problem that involves a mismatch between what people actually want and the target they choose, thereby decreasing their satisfaction.

To avoid the mismatch, one solution is to ask specialists (e.g., a concierge) to give advice on choice behavior, such as suggesting appropriate evaluation criteria for the decision or recommending items. The motivation of our research was to build a basis of interactive systems to assist users' choice behavior similar to a concierge. If an interactive system can assist users' selection behavior as an expert system, it helps users select targets [Häubl and Trifts, 2000, Chen et al., 2013].

One approach to support users' decision making is where an assistant system has a model of user preferences. The system can assist users' choice behavior by estimating user preferences based on the user model. Here, one limitation of the user-model based interactive assistance is that the effectiveness depends on the model construction. In this study, we focused on a user's choice situations, where

he or she chooses one item from several alternatives of *multi-attribute content*, a digital catalog whose displayed items have multiple attribute types. Here, we propose a framework of overcoming the limitation in a data-driven manner.

1.2 Users' Choice Behavior

Choice behavior is concerned with goal-directed behavior [Hansson, 1994]. The goal depends on each user and situation. Here, utility is known as one measure of goodness of the decision.

1.2.1 Utilities

Utility is a measure of preferences for alternatives, and it represents the satisfaction of consumers in economics. Because user preferences cannot be directly observed, a user preference is modeled by the utility function [Ehrgott et al., 2010] in a wide range of research fields, such as operations research. The value of a utility function represents the degree of satisfaction based on current preference. That is, when a person prefers item I_1 to I_2 , the utility of I_1 becomes higher than the utility of I_2 .

Exploitation/Exploration

Two kinds of choice behavior can be considered in terms of choice goals: exploitation and exploration [Athukorala et al., 2016, White et al., 2006]. Exploitation is a choice to select an alternative that matches the user preference; the goal is to increase the utility of the choice. In contrast, exploration is a choice where a user does not have an explicit preference; the goal is to increase his or her experience through the choice. Users sometimes switch these two kinds of choices and achieve better alternatives, which are known as a kind of multi-armed bandit problem. Although exploration is not the most appropriate choice regarding the specific preference, exploration has a potential to update user preferences, and as a result users can choose alternatives with higher utility.

1.2.2 Multiple Criteria Decision Analysis

Multiple criteria decision making (MCDM) is an analysis or a problem in which users make selections from many alternatives by evaluating multiple conflicting

criteria. Choosing alternatives among many candidates that have multiple attributes is one example of MCDM because users evaluate each value of an attribute. If only one criterion is important, users can easily choose items that have higher utility. However, if the number of criteria exceeds one, the decision becomes harder because the item having the highest utility may conflict regarding each criteria. In this way, because MCDM is a complex problem to achieve a goal conflicting with nature, several operations research techniques have been applied for robustness and rational decision making [Ehrgott et al., 2010].

Model of users' decision strategies

Existing studies proposed several models of decision strategies [Fishburn, 1970, McAllister et al., 1979, Peter and Tarpey, 1975, Tversky, 1972, Pohekar and Ramachandran, 2004]. One of the basic approaches is well known as a weighted additive model (WAD) and weighted sum model (WSM) [Fishburn, 1970], which is a mathematical model of utility and which gives weights for each of the criteria and calculates the utilities of the alternative using a weighted summation of utilities for the criteria. In the WAD, the utilities of item i are given as $S_i = \sum_r w_r u_r(i)$, where w_r is the weights of criterion r and where $u_r(i)$ is the utility of item i regarding criterion r . Another approach is a weighted product method [Pohekar and Ramachandran, 2004], which compares two alternatives by productivity. For example, the utilities of two alternatives i and j can be compared by $S_i/S_j = \prod_r (u_r(i)/u_r(j))^{w_r}$. However, these approaches require much effort because users need to give weights to all evaluation criteria and to calculate all utilities for the criteria. That is, users' choice behavior has some trade-off between the results and effort.

Therefore, users often choose alternatives using non-compensatory strategies such as elimination by aspects (EBA) [Tversky, 1972] or lexicographic (LEX). Users who adopt these strategies compare alternatives one criterion at a time. The LEX strategy is where users focus on only one criterion and choose an alternative that maximizes its utility regarding the criterion similar to single criterion decision making (i.e., $\arg \max_r u_r(i)$); this is one of the most simple decisions in MCDM. The EBA strategy is more complex and ordinal. For each time step, the users gradually eliminate the alternatives that do not satisfy that threshold by focusing on a criterion and by deciding one threshold to the value of utility regarding the criterion. Although these techniques do not simultaneously consider all evaluation criteria and are therefore non-compensatory, users can easily choose

Clearness of displayed information	Clear	Daily necessities Daily food Cloth	Electronic devices Special dinner Gift	Car House
	Unclear	Movie Book	Lecture	Job hunting
		Low	fair	High

The degree of impact (Life / Times)

Figure 1.1: Categorization of the domain of choice behavior

alternatives, ensuring less cognitive effort.

Decision strategies and types of selection targets

Users' choice behavior can be different based on content domain, so effective assistance must be able to adapt. Figure 1.1 shows an example of a map of content domains. We divide the domain by the impact on users' lives and the clearness of the displayed description of alternatives. The impact is approximately calculated based on the decision times.

The impact and clearness may be associated with the amount of effort, appropriate strategies, and needs for interactive assistance. If the impact of the decision is small, users decide the alternative at low cost; LEX or EBA may be appropriate for the decision. In contrast, if the impact is large, users may exert high effort, i.e., users can carefully evaluate alternatives from several evaluation criteria; the WAD or WSM may be appropriate for the decision. If the problem complexity increases, the difficulty of the decision making correspondingly increases, and the effectiveness of the interactive assistance increases. In fact, Dijksterhuis et al. said that the users consider only two or three criteria to choose alternatives by the WAD [Dijksterhuis and Nordgren, 2006].

Also, if the displayed information is clear enough to imagine the alternatives (e.g., the specs of electronic devices are almost completely described by the text), the users can choose alternatives by displaying information. However, if the provided information is unclear and does not directly describe the detail of the alternatives, interactive assistance is effective to support users imagining what the alternatives are.

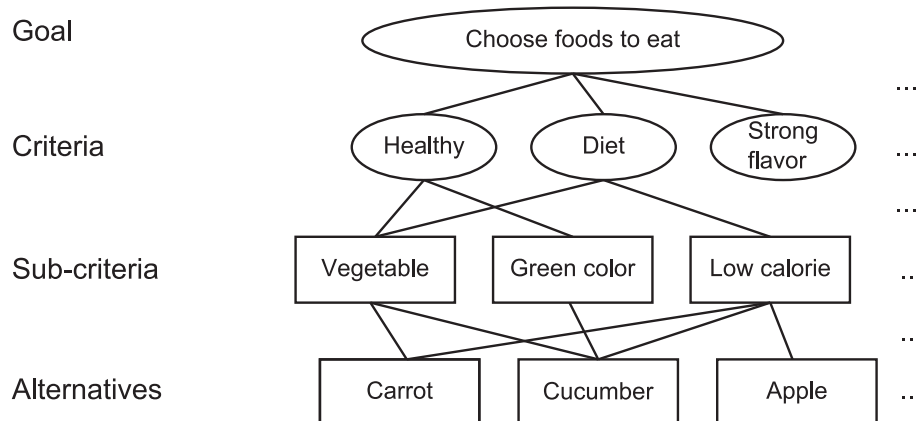


Figure 1.2: An example of structure of the AHP

1.2.3 Analytic Hierarchy Process

The analytic hierarchy process (AHP) [Saaty, 1980], which is a structured technique for organizing and analyzing user preferences, is one basic approach for MCDM as an extension of the WAD. In the AHP, the decision problem is modeled as a hierarchy that consists of the users' decision goals, evaluation criteria, and alternatives as shown in Figure 1.2. Those alternatives are evaluated according to each evaluation criterion.

The procedure of the basic AHP consists of the following four steps.

1. Structure the decision problem as a hierarchy containing the decision goal, criteria, sub-criteria, and alternatives.
2. Evaluate elements in the hierarchy using pairwise comparisons.
3. Calculate utilities of alternatives by weighting and adding these evaluations.
4. Decide one alternative based on the utilities and evaluations of these elements.

Here, step 1 is important for decision making because structuring the decision making enables the users to understand the problem [Bhushan and Rai, 2004].

The AHP is based on the WAD strategy, but AHP introduces the pairwise comparison to evaluate each element and to normalize the evaluation scores in each layer. These extensions provide users with a framework for organizing their weights on each of the elements.

1.2.4 User Preferences in Choice Behavior

User preferences consist of several factors. In several research fields including education, these preferences are called *individual interests* and *situational interests* [Rotgans and Schmidt, 2011, Wentzel and Wigfield, 2009]. Situational interests are interests that are elicited by the external factors such that a student takes interest in a historical character when a teacher mentions a historical event. In contrast, individual interests are users' own preferences. Each user has his or her own individual interests that naturally develop over the course of life, and a part of these interests are activated based on the current situation.

In this study, we assumed a situation where a decision problem can be structured a hierarchy such as the AHP. An assistant provides information that is a part of the structure to users; for example, the assistant displays part of the set of items as catalog content. This provided information to the users is referred to as the *situation*. In the structured problem, the users' interests can be considered for each of the elements. An assistant provides a situation to users, and users may take interest in the elements in the structure; we call them *activated interests*.

Because a decision problem is organized in a hierarchical structure, users' activated interests can be represented once by an importance weight of elements in the top level (criteria). In this study, we specifically focused on these activated interests of top level elements and name *selection interests*.

1.2.5 Phase of Choice Behavior

Users' choice behavior consists of several phases [Beach and Mitchell, 1978, Hansson, 1994, Russo and Leclerc, 1994]. These phases mean that users do not always compare items but browse content toward several sub-goals. In this section, we categorize these phases into two states: during choice and after choice.

During Choice

First, users recognize the problem surrounding the decision based on the current situation. Then, users adopt strategies for making the decision depending on the decision problem such as the impact. In this stage, users recognize the parameters of the problem and what the users need to know for the decision. After that, users gather the information on the alternatives and adopt the aforementioned various strategies to process information.

The last stage can be divided into two parts based on sub-goals: divergent and convergent [Kaner et al., 2011, Ehrgott et al., 2010]. These phases are especially known in group decision making. The divergent phase is to activate interests by updating the situation and thereby acquiring the information. The convergent phase is to eliminate alternatives based on current activated interests. This process of decision making is called the *diamond of participatory decision-making* especially in the field of group decision making.

After Choice

Users interact with chosen alternatives and obtain satisfaction if the alternatives have the same number of utilities or many more than users imagine when making decisions. At that time, if the utilities of chosen alternatives are less than what is imaged, that mismatch decreases user satisfaction [Dijksterhuis and Nordgren, 2006, Milliman and Decker, 1990]. This post-purchase dissonance tends to occur when the impact of decisions is large [Cohen and Goldberg, 1970]. In other words, to obtain satisfaction regarding the choice, we need to imagine the actual situation of the decision especially for high importance. That is, users need to know the content domain (i.e., whether or not it is divergent enough) and then to clear their selection interests (i.e., whether or not it is convergent enough).

1.3 Selection Assistance

As mentioned already, a user has a goal and adopts several strategies for the choice behavior. Here, several kinds of selection assistance exist to help users achieve their own goals. In fact, an assistant has its own intention such as the assistant wanting to sell a certain item. Therefore, the assistant's suggestion has some bias. In this study, we specifically focused on a situation where a user and an assistant exist and do not take account of an assistant's intention by assuming that the assistant is unbiased.

1.3.1 Component in Selection Assistance

The situation of selection assistance is shown in Figure 1.3. Two players, a user and an assistant, exist, and they each have their own preference, knowledge about the content domain, and an intention. The assistant also has estimated

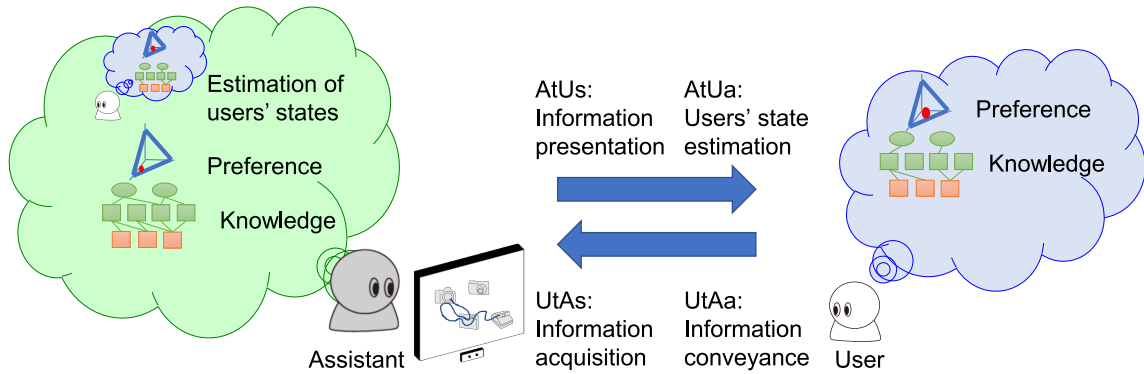


Figure 1.3: Component in selection assistance

user states, and the assistant facilitates user's selection behavior by using these states. Actions between the user and the assistant have two directions: assistant to user (AtU) and user to assistant (UtA). Also, these actions can be divided into situation-related actions (AtUs and UtAs) and activation-related actions (AtUa and UtAa).

1.3.2 Information Presentation to Users

An information presentation action features an assistant presenting a situation to users. This is the most basic assistance for the users' divergent sub-goal. Here, an effective presentation can vary depending on the content domain.

Modalities of Presentation

An assistant can present information to a user with several modalities. One of the main modalities is displaying the content because users obtain much more information from gazing. Also, the assistant uses a voice or gestures to make a user imagine the alternatives. In addition, the assistant often lets a user have a short experience such as sampling food.

1.3.3 Estimation of Users' Internal States

An assistant not only presents information but also estimates users' activated interest from users' reaction to the assistant's question or users' action. By estimating the users' activated interests in addition to his or her explicit commands or verbal actions, the assistant can adoptively change the way of assistance.

This estimation consists of not a single action but a sequence of actions. When

the assistant estimates the users' activated interests, it confirms the estimation by asking questions to the users. These sets of actions are also divided into estimations of users' internal states.

1.3.4 Information Acquisition to Assistants

Information acquisition is a user's proactive actions to update the situation. The user can obtain useful information from the assistant by not only reacting to an assistant's presentation but by proactively conveying the user's needs or preference to the assistant.

1.3.5 Information Conveyance to Assistants

Information conveyance is a reactive action by the users. By answering the questions from the assistant, the users tell the assistant their activated interests. Information conveyance is a part of the assistant's estimation because the assistant determines the assistant's estimation by the users' conveyance and updates the estimation. In contrast to the assistant's estimation, both the conveyance and acquisition are the users' intentional actions, so the users tell the assistants by explicit actions such as speech or clicking the answer.

1.3.6 Interaction During Assistance

Interactive assistance can be categorized by which information the assistant uses to present information to the user. That is, the sequence of AtU and UtA.

Non-Interactive Assistance: AtUs

This assistance only presents information to the users. Its aim is to give users knowledge about the decision problem. Examples of non-interactive assistance are paper catalogs and advertisements.

Interactive Assistance

Interactive assistance changes the displayed content based on the interaction, in contrast with non-interactive assistance.

Let us consider two kinds of interaction: reactive assistance and proactive assistance. Reactive assistance changes how information is presented based on

Table 1.1: Categorization of interactive assistance system

	Reactive	Proactive
Interaction construction	FindMe [Burke, 2002] Intensional Summaries [Polifroni and Walker, 2008]	
User model	MATCH [Walker et al., 2004]	Sightseeing guidance [Misu et al., 2011] Course adviser [Dodson et al., 2013]

users' explicit action, and proactive assistance changes that information based on the assistant's estimation of the user preferences.

Reactive Assistance: UtAs $\rightarrow$ AtUs Examples of reactive assistance are filtered search and faceted search. In these types of assistance, the assistant prepares several filters beforehand such as "more than a 4-star review" or "the category is a Japanese restaurant." The user chooses several filters to narrow down a list of alternatives. By reacting to users choosing filters, the assistant presents alternatives that satisfy those filters. This approach can interactively decrease candidates by eliminating alternatives that do not satisfy users' needs through a user-initiated interaction.

Proactive Assistance: AtUa $\rightarrow$ AtUs Proactive assistance is where the assistant helps users based on not only users' explicit actions but also based on the assistant's estimation of users' internal states. Examples of proactive assistance are a concierge or salesperson. First, they present information to users and probe users' internal states by observing the reaction of users. Then, they react to users such as by recommending alternatives that may match their preferences. Proactive assistance works well specifically when users do not have a clear understanding of what they want. In such case, reactive assistance cannot help users with their decision making because users have a hard time inputting appropriate information to the assistant.

1.3.7 Interactive Decision-Support System

In addition to the aforementioned categorization, designing interactive systems can be divided into two types, interaction construction and user model construction, as shown in Table 1.1.

The first approach is to construct an interaction so that the system updates providing information based on individual user feedback to currently displayed items [Burke, 2002, Polifroni and Walker, 2008].

The other approach is to build a user model so that the system provides appropriate information to users [Misu et al., 2011, Walker et al., 2004, Dodson et al., 2013]. This approach enables the system to generate a user-tailored response based on the users' internal states such as directly presenting alternatives that fit the users' needs. The advantage of introducing a user model is that the system is able not only to return a user-tailored response but also to proactively interact with users by estimating their internal states [Misu et al., 2011, Dodson et al., 2013] (i.e., system-initiated interaction). Therefore, the approach building a user model is suitable for a preference construction situation.

1.4 Gaze Behavior in Catalog Content Browsing

To estimate user preference as aforementioned, we specifically focus on the users' gaze behavior as one type of non-verbal behavior. Gaze behavior is considered to be a good indicator for estimating various aspects of users' decision-making [Orquin and Loose, 2013].

1.4.1 Features of Eye Movements

Basically, eye movements are captured as 2-D points on a screen. For gaze analysis, eye movements are then encoded using several techniques [Jacob and Karn, 2003].

Primitive motion of eye movements

First, we describe two kinds of primitive motion of eye movements while users browse static content.

Fixation A fixation or visual fixation is keeping a point of gaze. The duration length of fixation can vary from less than 100 ms to more than several seconds depending on the task, e.g., the mean fixation duration length during silent reading and during visual search are 225 ms and 275 ms, respectively [Rayner, 1998].

Saccade A saccade is very quick eye movements in between fixations. The velocities are as high as 500° per second. During saccades, input information is reduced; this phenomenon is called saccadic suppression.

Motion features of eye movements

The motion features of eye movements are the features of the aforementioned two primitive motions: fixation and saccade. Existing studies often utilize multiple features of these two motions, such as fixation duration, fixation count, and saccade length [Sugano et al., 2013, Eivazi and Bednarik, 2011, Bednarik et al., 2012, Cole et al., 2013]. For example, Sugano et al. used a combination of various fixation and saccade features and assessed the contributions of each feature for a classification task using a random forest algorithm [Sugano et al., 2013].

Incorporating the semantic information of browsed content is another approach. The semantic labels of content are manually annotated to the area of interest (AOI) beforehand, and gaze behavior is represented as a sequence of AOIs [Brandherm et al., 2008, Nakano and Ishii, 2010, Kübler et al., 2014]. AOIs are defined by each of the targets in the content [Brandherm et al., 2008] or part of the alternatives such as the “agent’s head” [Nakano and Ishii, 2010].

Semantic information on the AOIs can also be utilized. Schaffer et al. [Schaffer et al., 2016] focused on a design structure of catalog content; items in catalog content are grouped by certain attributes. They characterized the gaze feature by using the relationship between two items such as a parallel relationship or contrast relationships. This representation is based on both semantic information and users’ behavior.

1.4.2 The Factors of Eliciting Users’ Gaze Behavior

Users’ gaze behavior is affected not only by users’ internal states such as preference and intention but also the effect of external factors such as users tending to look at the center of the content [Borji, 2012, Judd et al., 2009].

External factors

Visual attention is stimulus driven and saliency based during content browsing, and the research of modeling visual attention has been an active topic for over 20 years [Borji and Itti, 2013]. The saliency map, proposed by Itti et

al. [Itti et al., 1998], is a representative implementation based on feature integration theory. Yonetani et al. [Yonetani et al., 2013] revealed that the spatial and temporal gap between the users' gaze and the dynamics of the salient region indicate the users' cognitive state without analyzing the semantic information of the displayed content.

Internal factors

The users' gaze behavior is affected by their internal factors from several aspects. Yarbus showed that the gaze behavior is strongly influenced by the tasks in which they are engaged or their cognitive goals [Yarbus, 1967]. Gaze behavior is used to estimate their variety of internal states including knowledge levels [Cole et al., 2013], cognitive abilities [Steichen et al., 2014], engagement [Ishii et al., 2013], and preferred items [Brandherm et al., 2008, Hirayama et al., 2010]. For example, Brandherm et al. [Brandherm et al., 2008] proposed an approach for estimating user's preferred target in displayed content by frequency and duration of gaze targets.

1.4.3 Users' Gaze Behavior and Decision Making

As mentioned, users adopt several strategies during decision making. Because users' gaze behavior also changes corresponding to the strategy, their decision state including strategy or phase can be estimated by observing gaze behavior [Ball, 1997, Shi et al., 2013]. For example, Ball [Ball, 1997] showed the relationships between users' decision strategies and a multiple-step transition of gaze targets in an information matrix consisting of items and attributes.

In contrast, gaze behavior also affects the users' decision making. Shimojo et al. [Shimojo et al., 2003] revealed that the users tend to choose alternatives that look at long durations.

1.5 Problem Statement

Let us assume an MCDM situation where a user chooses one item from several alternatives of *multi-attribute content*, a digital catalog whose displayed items have multiple attribute types. Moreover, the decision has a fair impact on the user's life, as shown in Figure 1.1, but the user does not have enough understanding of his or her own preference. The aim of this study was to build a basis of interactive

assistance for users' choice behavior in such situations. Specifically, we focus on users' gaze behavior as mentioned in Section 1.4, the system assists users' choice behavior based on the estimation of users' internal states from users' gaze behavior.

As mentioned, several kinds of goals can be considered during users' choice behavior. The goals of assistance that had to be taken into account in this study are as follows:

- I) An assistant helps the user recognize the structure of a decision problem by mentioning possible evaluation criteria (i.e., divergence of the users' knowledge).
- II) An assistant helps the user choose items that match the user's preference by estimating the user's selection interests and recommending items (i.e., convergence of the users' selection interests).

For the goal I), the system needs to have the structure of the problem beforehand. The prepared evaluation criteria limit the decision-support system because the effectiveness of the support strongly depends on the chosen evaluation criteria. However, preparing evaluation criteria themselves in a top-down manner is difficult in many cases because appropriate evaluation criteria can vary depending on content, individuals, and engaging tasks.

For goal II), the system needs to estimate users' selection interests from users' non-verbal behavior. Laddering [Reynolds and Gutman, 1988] is a well-known technique to probe a user's preference behind the target selection in marketing. Marketers change their sale strategies by repeatedly asking the reason for having chosen targets. In contrast to laddering, which explicitly asks the reasons to users verbally, the situation we assume is that users do not have a clear understanding of their preference, and inputting their preference to the systems explicitly is difficult.

Also, the system needs to evaluate alternatives regarding users' selection interests for goal II). To estimate users' selection interests, preparing evaluation criteria beforehand and estimating users' selection interests using these criteria can be one solution. Here, since the evaluation also depends on situations, we need to address how to represent evaluation criteria so that the system can associate evaluation criteria with alternatives.

To sum up, toward achieving the goals, we addressed the following issues: how to represent/prepare evaluation criteria; and how to estimate users' selection

interests from gaze behavior.

1.6 Contributions

The main contributions of this study are three-folds: (1) We introduce a novel representation of users' selection interests by introducing a notion of *aspects*, which represents why users look at (i.e., evaluation criteria); (2) we focus on users' comparison behavior that reflects users' aspect of focus and propose a comparison detection approach based on the bias of gaze behavior; and (3) we propose a generative model of gaze behavior to learn aspects and to estimate users' selection interests behind users' comparison behavior.

1.6.1 Representation of users' internal states

While some studies that estimated users' interest from gaze represented users' interests as an item of displayed content in which the user takes interest [Brandherm et al., 2008, Hirayama et al., 2010], these representations and estimation are not enough for recommendations. Instead, the reason a user chooses a target (i.e., evaluation criteria) is important to recommend appropriate alternatives similar to the ladder. In this study, we focus on *aspects* of items as possible reasons for decision making and by characterizing these aspects of items by the degree of association with attribute values, we propose a hierarchical structure of decision problem. Here, users' selection interests are referred to as the importance weights of these aspects because the importance weights of the aspects can be an indicator on which ones the user focuses for the decision as mentioned in Section 1.2.4.

1.6.2 Users' comparison behavior

Specifically when users browse a multi-attribute content, it is difficult to know why the users look at only by observing the users' gaze targets. Users often fix a certain criterion and compare targets that satisfy the criterion during the comparison, so users' comparison behavior is expected to reflect evaluation criteria. Here, one observation of users' comparison behavior is gaze behavior (i.e., transition of the gaze targets), and gaze moves without users' explicit intention and reflects users' implicit intention. This is why we focused on users' gaze behavior in this study.

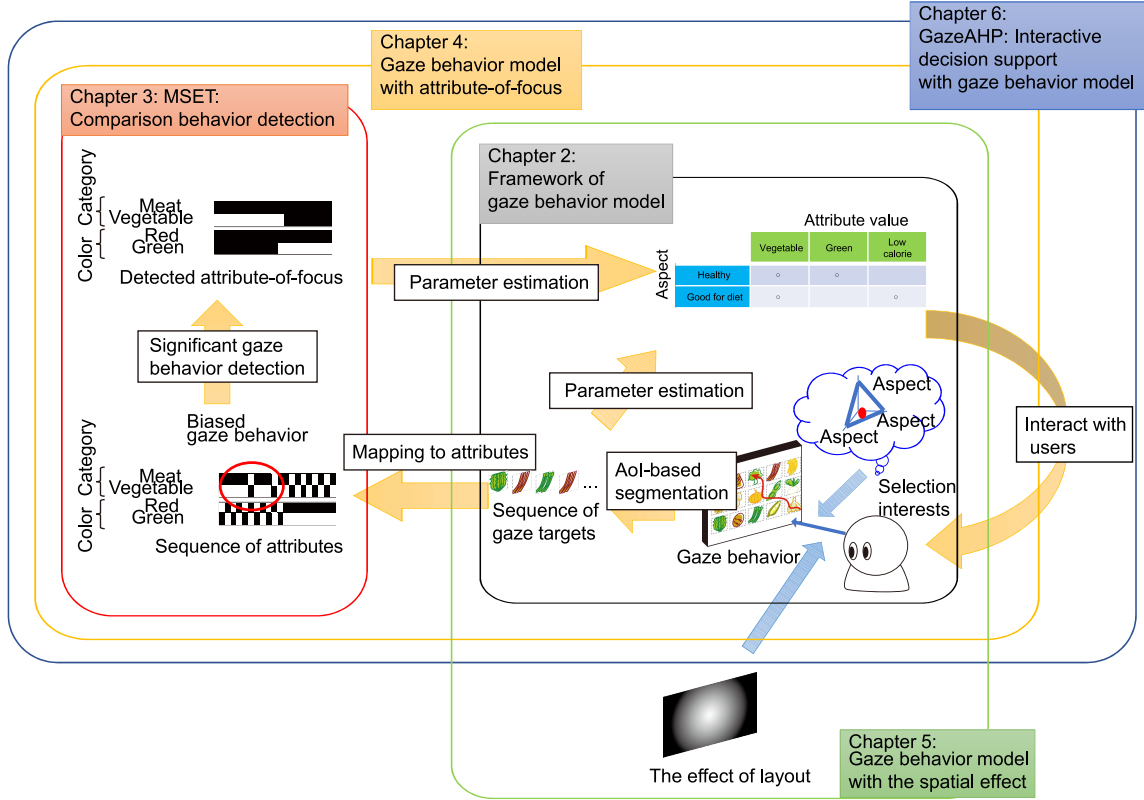


Figure 1.4: Structure of the thesis

1.6.3 Gaze Behavior Modeling

As mentioned in the Section 1.5, preparing evaluation criteria in a top-down manner is difficult in many cases. One approach overcomes this difficulty by using a data-driven method and by finding possible evaluation criteria from users' behavior [Jin et al., 2004, Iwata et al., 2009, Bhargava et al., 2015]. We herein borrow the idea of topic models [Hofmann, 1999] and propose a generative process of gaze behavior using aspects. By observing a much amount of users' gaze behavior, this gaze behavior model learns aspects in a data-driven approach and estimates selection interests based on learned aspects.

1.7 Structure of the Thesis

Figure 1.4 shows the structure of the thesis. We first describe a framework of the proposed aspect-oriented gaze behavior model in Chapter 2. By introducing this probabilistic generative model of gaze behavior, the proposed framework estimates users' selection interests from users' gaze behavior based on probabilistic

inference. In addition, once selection interests are estimated, because the proposed model is generative, the users' items of interests or region of interests can be predicted.

In this thesis, we specifically focus on two difficulties, which are important for gaze behavior modeling, in the following chapters. (1) Multiscale comparison behavior detection is proposed in Chapter 3 to overcome the difficulties described in the previous section. In Chapter 4, the gaze behavior model is extended by taking attributes-of-focus into account, and (2) the effect of the layout is taken into account in Chapter 5. Once aspects are learned, an assistant can support users' decision making by estimating their preference based on those aspects. We show aspect-based interactive assistance in Chapter 6.

Chapter 2: Framework of Aspect-Oriented Gaze Behavior Model

In Chapter 2, we introduce a notion of *aspects* and a framework of gaze behavior model to learn the aspects and to estimate selection interests from gaze behavior. The proposed framework is based on the topic model, and by modeling a generative process of gaze behavior stochastically, aspect learning and selection interest estimation are achieved using the hidden parameter estimation. Once aspects are learned from a large amount of users' gaze behavior (i.e., in a data-driven fashion), their selection interests can be estimated.

Chapter 3: Multiscale Exact Test: Comparison behavior detection using statistical testing

In Chapters 3, 4, we introduce a framework of aspect learning by taking users' comparison behavior into consideration. To do that, we first utilize a multiscale comparison behavior detection approach named the multiscale exact test (MSET) in Chapter 3.

The MSET is a method to detect users' attribute values of focus, which we refer to as *attributes-of-focus*, from gaze behavior. Because users' comparison behavior can be observed temporally and partially, short-term analysis is applied to detect attributes-of-focus for gaze behavior. An appropriate window size is also determined to introduce short-term analysis; we therefore apply a null hypothesis test for multiple scales. By modeling users' neutral browsing behavior, this test identifies the periods of users' significant gaze behavior and detects attributes-of-focus from identified periods.

Chapter 4: Gaze behavior model with attributes-of-focus for aspect learning

Then, we introduce a framework of aspect learning with attributes-of-focus in Chapter 4. In this chapter, we specifically focus on users' attributes-of-focus, which can be observed temporally and partially from users' gaze behavior (i.e., a part of a sequence of attribute values of gaze targets). The proposed framework first detects users' attributes-of-focus using the MSET introduced in Chapter 3 and deals with the detected attributes-of-focus as an observation of a gaze behavior model. This two-step approach enables the gaze behavior model to learn aspects by explicitly taking the users' attention resource into account.

Chapter 5: Gaze behavior model with spatial effect for item recommendation

In Chapter 5, we focus on the spatial effect on the gaze behavior. As we mentioned in Section 2.6.1, the users' gaze behavior is affected both by internal and external factors. Therefore, how to distinguish these two sets of factors is the question to estimate users' selection interests from gaze behavior.

Our approach to address the issue is to deal simultaneously with these two factors in a unified generative model. Here, we introduce an extension of gaze behavior model by taking account of the effect of the layout.

Chapter 6: GazeAHP: Tracing selection interests and supporting users to recognize the structure of problems

Finally, in Chapter 6, we show aspect-based interactive assistance by utilizing the MSET introduced in Chapter 3 and selection interests estimation introduced in Chapter 4. The interactive assistance is based on AHP and named *GazeAHP*. Here, we specifically focus on two functions of assistance, probing and sorting, and show the effectiveness of our proposed approach from several viewpoints in interactive assistance.

Chapter 2

Framework of Aspect-Oriented Gaze Behavior Model

2.1 Introduction

One of the important issue to estimate users' selection interests is how to construct the state space of interests. Some existing studies represented users' interests as on which targets the users focus [Brandherm et al., 2008, Hirayama et al., 2010, Jawaheer et al., 2014]. However, this representation is not enough when it comes to building a recommender system because this representation is only for a displayed content and has no information about not displayed alternatives.

To address this issue, we introduce *aspects* of items which can be reasons why users look at alternatives (i.e., evaluation criteria). Since these aspects can vary depending on users and situations, it is hard to prepare in a top-down manner. One approach for overcoming this difficulty is to learn aspects from users' behavior, that is, in a data-driven fashion.

Among a variety of observable behaviors during users' content browsing, users' comparison behavior is one of the most important cues to learn evaluation criteria. This is because users tend to fix an evaluation criterion of focus for a certain period, and alternatives that users compare may therefore be related to the evaluation criterion of focus. In particular, gaze behavior conveys rich information to trace users' comparison behavior because users compare several items in a short term by gaze, in which the evaluation criteria of focus can be assumed to be fixed. Besides, eye movements are affected by users' implicit intention. Therefore, the evaluation criteria can be learned without users' explicit command.

In this chapter, we propose a probabilistic aspect-oriented gaze behavior

model to learn aspects, estimate users' selection interests and predict items of interests. The proposed model is an extension of the probabilistic Latent Semantic Analysis (pLSA) [Hofmann, 1999], one of the topic models, to include the characteristics of content browsing as described in Chapter 1.

The contributions of this chapter are as follows:

- We introduce a notion of *aspects* for users' evaluation criteria of decision making and introduce a novel representation of users' selection interests.
- We propose a framework of generative gaze behavior model to learn aspects and to estimate users' selection interests from gaze behavior.

2.2 Obtaining State Space of Users' Internal States from Users' Behavior

One of the popular issues in the analysis of human-computer interaction is the estimation of latent user states from observed users' behavior (e.g., eye movements), where the states include interests [Brandherm et al., 2008, Hirayama et al., 2010], intentions [Eivazi and Bednarik, 2011, Bednarik et al., 2012, Ishikawa et al., 2012] and attentive states [Yonetani et al., 2012]. The underlying approach of these studies is to extract various gaze features, such as fixation duration and saccade length, and to associate them with discrete user state labels (e.g., interested in item A, B, C, ..., high or low concentration) in a supervised learning fashion. Using these approaches enable us to estimate mental states from newly observed gaze data. However, these approaches need to assume what kind of states users are likely to become and give the labels of states in a top-down manner beforehand. This limitation is critical for our situation since a variety of aspects can be considered for users' interests in a general situation.

On the other hand, data-driven approaches to learn mental states from observed eye movements are proposed in the fields of information retrieval (ranking) [Puolamäki et al., 2008, Pasupa et al., 2009, Sugano et al., 2013], recommendation [Yoshitaka et al., 2007], and interaction mining [Jayagopi et al., 2012]. However, these approaches deal with users' interests toward only displayed items [Sugano et al., 2013, Yoshitaka et al., 2007] (or interaction targets [Jayagopi et al., 2012]). In addition, the effect of layout design cannot be separated from gaze of item region [Puolamäki et al., 2008, Pasupa et al., 2009].

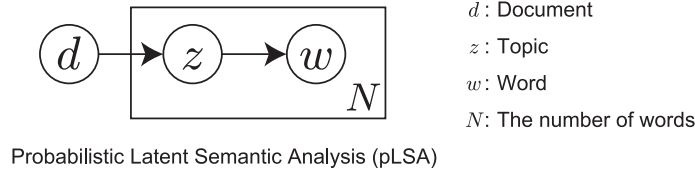


Figure 2.1: Graphical model of the pLSA

2.2.1 Learning State Space by Topic Models

Topic models such as the pLSA [Hofmann, 1999] and the Latent Dirichlet Allocation (LDA) [Blei et al., 2003] are the typical approach to obtain several bases that well represent observations in a data-driven fashion. The concept of the topic models is that by stochastically modeling a generative process of observation, the bases of observations are learned as a latent parameter estimation.

Topics or trends can also be learned from not only semantic information of content but also users' behavior [Hofmann, 2004, Jin et al., 2004, Iwata et al., 2009]. In this case, these learned topics can be regarded as evaluation criteria, i.e., the bases of the users' internal state space to estimate their interests.

We here show an example of obtaining the bases of observation by topic models using the pLSA. In the pLSA, topics are considered as the bases of the state space of words, and a topic z is characterized by a categorical probability distribution of an observation $\{P(w|z)\}_w$, where w is a word. Also, a document d , which consists of words, is represented as a mixture of topics z by a categorical probability distribution $\{P(z|d)\}_z$.

The generative process of words in the pLSA is as follows. For each document d , a topic z is first determined according to the probability distribution $\{P(z|d)\}_z$. Then, a certain word w is generated according to the probability distribution of an observation $\{P(w|z)\}_w$, conditioned by topic z . By iterating these process N times, the observations of a document are finally generated as a sequence of N words. The graphical model of the pLSA is shown in Figure 2.1. Each node shows one parameter and the edge of this graphical model shows a causality of parameters. The plate in Figure 2.1 means that the process on the plate iterates several times. With this generative model, suppose a large number of documents are observed, multinomial parameters $\{P(w|z)\}_w$, which characterize topics, can be learned in a data-driven fashion.

2.3 Learning Evaluation Criteria from Gaze Behavior

As mentioned in the previous section, some existing studies learn users' internal states including interests behind users' purchase behavior, i.e., decision behavior. While these techniques represent users' state space by alternatives (e.g., learned trends consist of a set of alternatives [Iwata et al., 2009]), the reason why the user chooses a certain alternative may not be because what the alternative is but because the alternative has a specific criterion. That is, alternative itself is not important but features of alternative become important for obtaining evaluation criteria behind users' choice behavior. In fact, laddering [Reynolds and Gutman, 1988], which is a well-known technique to probe users' latent intention behind purchasing behavior in the field of marketing research, repeatedly asks users "why is that important to you?" to understand reasons to choose the alternative behind features of the alternative. Since the situation we assume is that users do not have a clear understanding of their preference in contrast to laddering, which explicitly asks the reasons to users verbally, in this thesis, we try to obtain evaluation criteria behind users' choice behavior from users' gaze behavior.

Here, the problem is that it is hard to know users' evaluation criteria behind users' gaze target specifically when each of alternatives has multiple attributes. To overcome this difficulty, we focus on "users' comparison behavior." As mentioned above, when users have evaluation criteria to compare alternatives, gaze targets may commonly satisfy that certain evaluation criteria. Therefore, observing users' comparison behavior can be a key to learning evaluation criteria. The approach can be applied to the situation where users browse several alternatives such as catalog content.

2.4 Component of the Gaze Behavior Model

In this section, we first introduce a notion of aspects and representation of users' internal states, which are component of the proposed model.

2.4.1 Aspects

In this study, we focus on "why the user compares the items" and introduce *aspects* of items as the possible reasons of comparison (i.e., evaluation criteria). Here, an "aspect of item" is a viewpoint those items are looked at, and we as-

2.4. COMPONENT OF THE GAZE BEHAVIOR MODEL

		Attribute value					Aspect	
		Vegetable	Green	Low calorie			Healthy	Good for diet
Item	Carrot	○		○	Item	Carrot	△	○
	Cucumber	○	○	○		Cucumber	○	○
	Apple			○		Apple	△	△
	Greenpeace	○	○			Greenpeace	○	△
		Attribute value						
		Vegetable	Green	Low calorie				
Aspect	Healthy	○	○					
	Good for diet	○		○				

Figure 2.2: An example of aspects

sume each aspect can be characterized by the degree of association with each of attribute values. For example, an aspect “healthy” is highly related to “vegetable” and “green color,” and “good for diet” is related to “vegetable” and “low calorie.”

Figure 2.2 shows an example of correlation between items, attribute values, and aspects. We assume that items can be seen from several aspects such as healthy food and good for diet because the item has attributes related to several aspects.

2.4.2 Representation of Users’ Internal States

In this study, the evaluation criteria are assumed to be broader perspectives than attribute values and can be directly evaluated from the decision goal. Here, we assume that users evaluate alternatives by *aspects* of items as mentioned above. While attribute values itself can also be considered as criteria, in this study, we regard attribute values as sub-criteria and explicitly distinguish attribute values from evaluation criteria.

By using aspects, users’ internal states can be represented as Figure 2.3. Specifically, users’ selection interests are represented by the degree of importance on the aspects. By giving the relationship in green area in Figure 2.3 from the knowledge base of the content domain, aspects can be learned as the degree of association with attribute values (blue area in Figure 2.3) and users’ selection interests can be estimated as importance weight on aspects (red area in Figure 2.3).

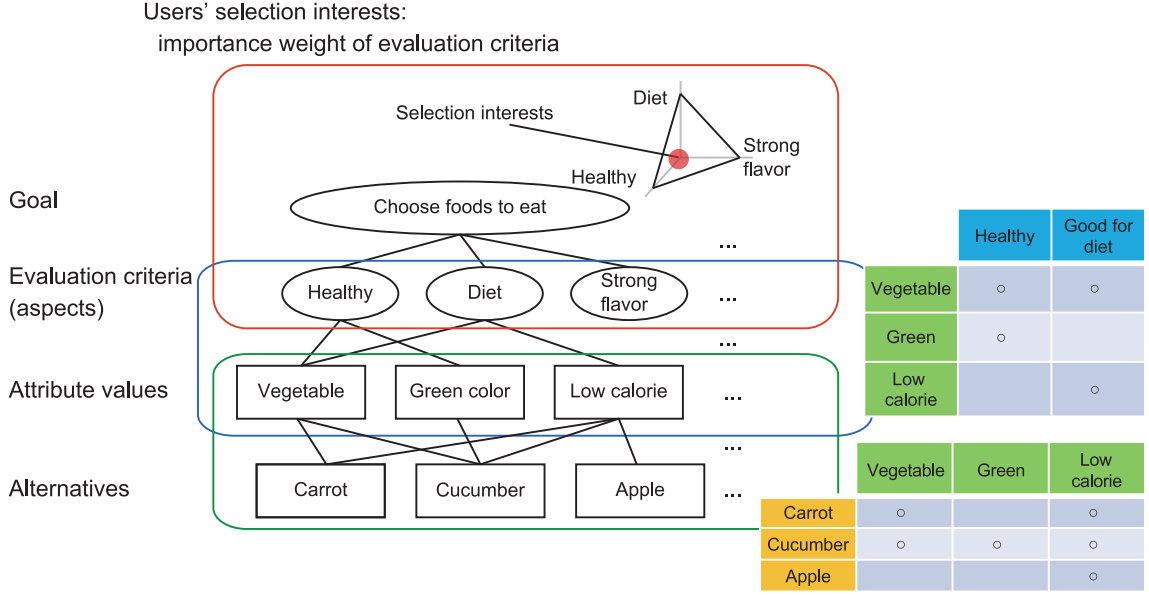


Figure 2.3: Hierarchical structure for decision making. Users' selection interests are represented as each evaluation criterion's weight of importance, and evaluation criteria are characterized by how they are related to each attribute value. Attribute values and alternatives are connected if alternatives have the specific attribute value.

2.5 Formulations of the Content-Browsing Situation

We here formulate users' internal states during content browsing. We assume a situation in which a user browses a content and denote this situation with mathematical notation as follows. The content displays a set of items $\mathcal{I} = \{I_1, \dots, I_N\}$, and every item are occupied a region $R_n \in \mathcal{R} = \{R_1, \dots, R_N\}$. Here, each region R_n is corresponded to an item I_n and each item occupies one item region. We here denote mapping function $I(r)$ to represent an item that occupies region r .

Every item has common M attribute types and takes one attribute value for each attribute type. Here, all attribute values as $\{V_1^{(1)}, \dots, V_{K_M}^{(M)}\} = \{V_1, \dots, V_K\}$ with the number of all attribute values $K = \sum_m K_m$, where K_m is the number of possible attribute values of attribute m , for convenience.

2.5.1 Model of Aspects

Let us assume a set of Q aspects $\mathcal{C} = \{C_1, \dots, C_Q\}$, which depends on a content domain but does not depend on users and only displayed items in the content. Here, we characterize these aspects according to the degree of association with

2.6. GENERATIVE PROCESS OF GAZE BEHAVIOR

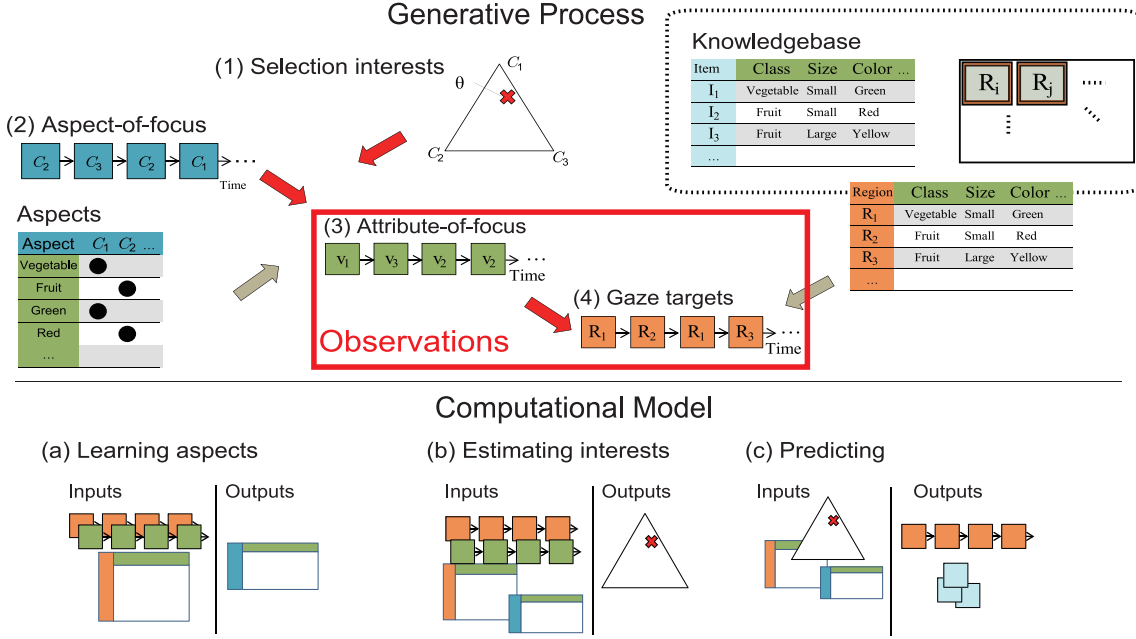


Figure 2.4: Generative process of gaze behavior from users' selection interests

each of attribute values. That is, we model an aspect C_q by a multinomial parameter $\mathbf{p}_q = (p_{q,1}, \dots, p_{q,K})$ ($p_{q,k} \geq 0, \sum_k p_{q,k} = 1$), where $p_{q,k}$ denotes how much the aspect C_q is related to attribute value V_k .

2.5.2 Model of Selection Interests

We represent users' selection interests by the importance weight of each aspect (i.e., evaluation criterion). Specifically, we model users' selection interests as an Q -dimensional parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_Q)$ ($\theta_q \geq 0, \sum_r \theta_q = 1$), where Q is the same as the number of evaluation criteria. This $\boldsymbol{\theta}$ determines a probability distribution that models how much the user focuses on each evaluation criterion, as will be described in the next section. Let us consider each browsing behavior for one decision making as a session, and users' selection interests are constant in each session for simplicity.

2.6 Generative Process of Gaze Behavior

We model a process of content browsing as follows. During session $s \in \mathcal{S}$, a user is assumed to have an interest modeled by $\boldsymbol{\theta}(s)$, and the user first decides aspect-of-focus $c_t \in \mathcal{C}$ at each time t , where c_t is determined according to the probability

distribution $P(c_t = C_q|s) = \theta_q(s)$. Note that, we assume $\theta(s)$ does not change during s for simplicity. Then, attribute-of-focus v_t is determined according to the probability distribution $P(v_t|c_t = C_q; \mathbf{p}_q)$ based on multinomial parameter $\mathbf{p}_q$, which represents a degree of association between attribute values and aspects. Finally, region $r_t \in \mathcal{R}$ is looked at depending on displayed item with the attribute-of-focus v_t .

In this study, not only the gaze targets itself but also attribute values of focus are utilized to observations of the model in order to take several factors into consideration as will described in Section 2.6.1. Particularly, in Chapter 4, the observations are attributes-of-focus, which are attribute values on which users specifically focus at a specific moment to directly take attribute values of focus into account. In Chapter 5, the observation is the regions that the gaze target allocates to take the effect of the spatial layout into account. Therefore, in this chapter, an observation at time t in session s is denoted as o_t^s and a set of observations are denoted as $\mathcal{O} = \{\{o_t^s\}_t\}_s$. Note that, the probability of an observation o_t is generated from an aspect-of-focus c_t is described as $P(o_t|c_t = C_q; \mathbf{p}_q)$ for simplicity although this probability is conditioned by not only the multinomial parameter of aspects but also the knowledge base of the content and the effect of layout when the gaze targets are considered as the observation.

Here, we assume that the relation between aspects and attribute values does not depend on a session. Then, the joint probability of an observation and aspect-of-focus for given session s is derived as follows:

$$P(o_t^s, c_t = C_q|s; \mathbf{p}_q) = P(o_t|C_q; \mathbf{p}_q)\theta_q(s). \quad (2.1)$$

Hence, the probability of observation o_t^s is given by

$$\begin{aligned} P(o_t^s|s; \{\mathbf{p}_q\}) &= \sum_{q=1}^Q P(o_t^s, C_q|s; \mathbf{p}_q) \\ &= \sum_{q=1}^Q P(o_t^s|C_q; \mathbf{p}_q)\theta_q(s) \end{aligned} \quad (2.2)$$

by marginalization over C in Equation (2.1). As a result, the probability of the sequence of observations in session s , $\{o_t^s\}$, is derived as

$$P(\{o_t^s\}_t|s) = \prod_{t=1}^{T_s} P(o_t^s|s; \{\mathbf{p}_q\}) = \prod_{t=1}^{T_s} \left\{ \sum_{q=1}^Q P(o_t^s|C_q; \mathbf{p}_q)\theta_q(s) \right\}. \quad (2.3)$$

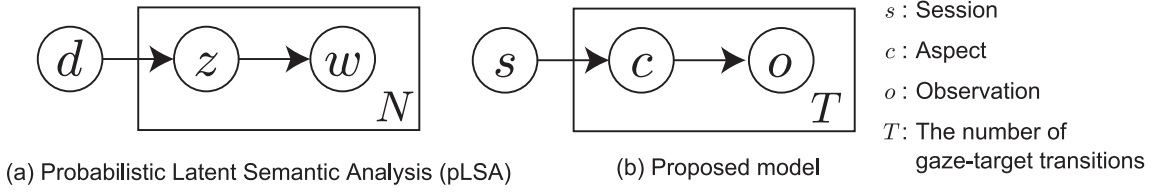


Figure 2.5: Graphical model of the pLSA and the proposed model

Figure 2.5 shows a relationship between the pLSA and the proposed model. Similar to topic learning in the topic models, the proposed model can learn aspects as the degree of association with attribute values.

2.6.1 Observation model

As mentioned in the previous section, this study deals with two kinds of observations to take different factors into account. This section represents observations of the gaze behavior model.

Gaze behavior segmentation

At first, we describe how to analyze users' gaze behavior. As we mentioned in Section 1.4, users' eye movements are observed as the 2-D points on a screen, and several kinds of annotations of eye movements are utilized to eye movements analysis such as sampling rate based analysis [Eivazi and Bednarik, 2011, Bednarik et al., 2012, Sugano et al., 2013] or AOI based analysis [Nakano and Ishii, 2010, Kübler et al., 2014]. Although duration length of gaze targets convey information, in this study, we specifically focus on the actions that users change browsed target and utilize AOI based analysis. This is because if several browsed target commonly satisfies a feature even when users change browsed target, the common features can be considered as the basis of users' comparison behavior. That is, the common features in users browsing behavior can be considered as evaluation criterion of the decision making.

For the AOI based analysis, we first annotate item regions on the content and map point of gaze into the gaze targets, and then define time t based on the transition of gaze targets. When users' gaze targets are determined, the region where gaze stayed below a threshold (under 100ms) are eliminated because users' fixation duration length to obtain information of targets is more than 100ms [Rayner, 1998] and shorter intervals of gaze are considered as measurement noise or temporarily passed region when the user changes gaze targets.

This sequence of gaze target is utilized to original gaze behavior in this study and we refer to the distribution of gaze targets in the sequence as *gaze distribution*.

Attribute values of focus

Attribute values on which users focus convey information of users' choice behavior [Ball, 1997, Pfeiffer et al., 2015]. For example, Ball [Ball, 1997] revealed that users' decision strategy could be estimated from the multi-step of transitions in the item-attribute matrix. As such, if users' attribute values of focus can be traced, users' aspect of focus (i.e., evaluation criteria) behind users' gaze behavior may be determined. Therefore, our approach is to adopt attribute values of focus as the observations of the gaze behavior model to obtain aspects from users' gaze behavior. However, attribute values on which users focus are not always directly observable because not all attributes are not displayed or the accuracy of eye tracking is low. To overcome the difficulty, we introduce a comparison behavior detection approach in Chapter 3, and apply the attribute-of-focus (i.e., detected attribute values of focus) to the observations of the gaze behavior model in Chapter 4.

Gaze target region

Users' eye movements are affected both users' internal state and external factor as mentioned in Section . That is, users' gaze target is determined by these both factors and therefore distinguishing these two effects are needed to trace users' selection interests. In particular, we specifically focus on the effect of absolute positions in the content (i.e., center bias [Judd et al., 2009, Borji, 2012]). To take external factors into account, in Chapter 5, not gaze targets but the regions that display gaze targets are dealt with observations of the model.

2.7 Estimation of Model Parameters

In this section, we describe how to estimate model parameters shown in Figure 2.4.

The parameters to be estimated are the interests in each session, $\theta(s)$, and the multinomial parameters that represent the degree of association between aspects and attribute values, $\{p_q\}$. That is, we estimate aspects themselves in the learning phase so that they represent training data well.

2.7. ESTIMATION OF MODEL PARAMETERS

Here, the probability of the sequence of the observations in session s is given by Equation (2.3). Therefore, the probability of the set of observations $\mathcal{O}$ is given by

$$P(\mathcal{O}) = \prod_{s=1}^S \prod_{t=1}^{T_s} \sum_{q=1}^Q P(o_t^s | C_q; \mathbf{p}_q) \theta_q(s) \quad (2.4)$$

In this thesis, we learn every parameter by solving optimize problem:

$$\begin{aligned} & \underset{\{\boldsymbol{\theta}(s)\}_s, \{\mathbf{p}_q\}_q}{\text{maximize}} && \sum_{s=1}^S \sum_{t=1}^{T_s} \log \left(\sum_{q=1}^Q P(o_t^s | C_q; \mathbf{p}_q) \theta_q(s) \right) \\ & \text{subject to} && \sum_q \theta_q(s) = 1 \quad \forall s \\ & && \sum_k p_{q,k} = 1 \quad \forall q. \end{aligned} \quad (2.5)$$

2.7.1 Learning Aspects (Figure 2.4 (a))

When we solve the optimization problem (Equation (2.5)) to learn aspects, input and output parameters of the algorithm are as follows:

Input parameters $\mathcal{O}$ (the set of observations)

Output parameters $\{\mathbf{p}_q\}$ (the degree of association between aspects and attribute values), $\boldsymbol{\theta}(s)$ (selection interests in each session)

To learn these parameters, we use the EM algorithm [Dempster et al., 1977] that repeats the following E-step and M-step until log-likelihood L converges. We give random values that satisfy constraint equations as initial values.

E-step computes posterior probabilities of latent variables, $\{P(C_q | s, o_t^s)\}_{s,t,q}$, under current parameters.

M-step updates parameters $\{\mathbf{p}_q\}$ and $\{\boldsymbol{\theta}(s)\}$ to maximize the log-likelihood under the obtained probabilities in E-step.

2.7.2 Estimating Selection Interests (Figure 2.4 (b))

When we estimate interests $\boldsymbol{\theta}(\hat{s})$ of new session $\hat{s}$, the parameters are as follows:

Input parameters $\{\mathbf{p}_q\}$ (the degree of association between aspects and attribute values), $\{o_t^{\hat{s}}\}$ (the sequence of observations of the new session)

Output parameters $\theta(\hat{s})$ (selection interests of the new session)

We also use the EM algorithm to learn the parameters. Note that, in contrast to learning aspects, we maximize only $\theta(\hat{s})$.

2.7.3 Predicting of Observation (Figure 2.4 (c))

When we predict an observation of this generative model from learned aspects and estimated selection interests, the parameters are as follows:

Input parameters $\{p_q\}$ (the degree of association between aspects and attribute values), $\theta(s)$ (selection interests of the session)

Output parameters $\{P(o_t|s)\}$ (probability that observation is generated)

These parameters are derived by Equation (2.2).

2.8 Extension of the Propose Model

In this chapter, we introduce the gaze behavior model based on the pLSA. Here, one possible extension to the model is to take a prior distribution of users' selection interests. Probability of θ is considered instead of the point-estimation of the θ , and the prior of the probability distribution $P(\theta)$ is modeled. LDA [Blei et al., 2003] is an extension of the pLSA by introducing Dirichlet distribution as a prior distribution of the topics. This extension enables us to represent users' selection interests with its ambiguity by a parameter of the distribution. In Chapter 6, we introduce a Dirichlet distribution to the model as a prior distribution and propose an approach to trace temporal changes of users' selection interests with its ambiguity.

Chapter 3

Multiscale Exact Test: Comparison Behavior Detection using Statistical Testing

3.1 Introduction

Evaluation criteria on which users focus gives a clue to assist users' decision making similar to the laddering [Reynolds and Gutman, 1988]. However, users sometimes only have a fuzzy understanding of their selection interests and cannot tell assistants about their evaluation criterion of focus. Therefore, users' evaluation criteria of focus estimation is a key technique to decision support. To do that, users' comparison behavior is essential as mentioned in Section 1.5. Specifically, this study focuses on users' gaze behavior because users' comparison behavior can be observed as the sequence of gaze targets.

From preliminary experiments, there are two observations:

- Users do not always compare several items and sometimes just browse content to obtain information on the displayed items.
- Users do not always take account of all attribute types of items but focus on some of them, and that attribute types of focus can change dynamically.

These observations indicate that users selectively switch their attention within an *attention resource*, a limited capacity of resources for attention. That is, both “when” the user compares needs to be identified and “which attribute types/values” of items reflect users' interests needs to be detected. We here in-

introduce a two-step approach to detect users' comparison behavior named Multiscale Exact Test (MSET). At first, by modeling a users' neutral browsing behavior, a hypothesis statistical test is adopted to identify users' biased gaze behavior. Then, comparing the histogram of gaze targets in the detected periods and neutral browsing model enable us to detect attribute values on which users' specifically focus, we refer to as *attribute-of-focus*.

The contribution of this chapter is as follows: We propose a multiscale comparison detection approach to identify users' attribute-of-focus while users browse multi-attribute content by detecting the bias of users' gaze behavior with its scale by the statistical test.

3.2 The Concept of the Multiscale Exact Test

As introduced in Section 3.1, to take account of users' attribute-of-focus while browsing multi-attribute content is crucial to obtain evaluation criteria. In this section, we propose an attribute-of-focus detection method from users' gaze behavior. The detected attribute-of-focus are used as inputs of the learning method of evaluation criteria described in Chapter 4.

To discover users' attribute-of-focus, we need to address (1) how to perform short-term analysis for attribute-of-focus detection and (2) how to determine the analysis window size. The reason why we need to deal with (1) is as follows. As mentioned in Section 3.1, users do not always compare items, and the users' comparison behavior can be observed temporarily. Moreover, even if the user focuses on certain evaluation criteria, the users' attribute-of-focus can be changed dynamically. For example, if a user focuses on "healthy" food, the attribute value of focus can change from "vegetable" to "green color." Therefore, to trace the change of users' attribute-of-focus, a short-term analysis for gaze behavior needs to be applied. Counting how many times each attribute value is looked at is not enough because the difference of the number of items having the same attribute value in the content is not taken into account. For example, if a catalog content contains five items and four of five are "green," the number of times a user looks at "green" items increases even if the user does not focus on any colors. Also, for (2), how long users pay attention to an attribute-of-focus depends on users and situations, and the appropriate analysis window size can be different. Therefore, to determine the analysis window size in a top-down manner is difficult.

To address these issues, we focus on a bias of users' gaze behavior in a short-

3.2. THE CONCEPT OF THE MULTISCALE EXACT TEST

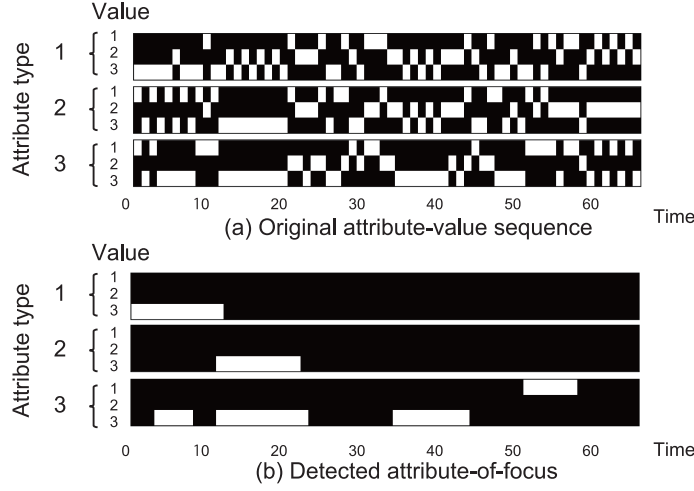


Figure 3.1: An example of detected attribute-of-focus. Each row shows an attribute value. Each column shows one attribute-value vector of browsed target, and white color denotes an item has that attribute value. (a) shows an original attribute-value sequence, (b) shows a sequence of attribute-of-focus detected by MSET.

term window. When a user takes an interest in specific evaluation criteria, the user tends to browse items that match certain evaluation criteria, and the users' gaze behavior has some bias. Our approach is to detect that bias from neutral browsing by a framework of the hypothesis test. Modeling neutral browsing with a simple probabilistic model and using "the user is browsing in a neutral manner in the time window" as the null hypothesis, the bias can be detected as a p -value (i.e., the probability that the observed behavior within a window obeys the neutral browsing model). To cope with the small sample size in a short-term window, we adopt an "exact test," which can be applied to different scales of windows including sizes with a few samples. As a result, the proposed method can be seen as a multi-scale analysis, and biases of users' gaze behavior are identified together with those window sizes. By observing the count of attribute values in the identified window, the attribute values on which the user focuses are also detected. Figure 3.1 shows an example of attribute-of-focus detection. (a) shows the original gaze behavior, i.e., the sequence of attribute values, and (b) shows the sequence of attribute-of-focus detected by the MSET. As shown in Figure 3.1, the MSET highlights the attribute values on which the user focuses.

In this chapter, attribute-of-focus are detected independently for each of attribute types in contrast to all attribute values are uniformly considered as observation in the previous chapter. Therefore, we redefine a notation of attribute

values. A set of possible values of the m -th attribute type is denoted as $\{V_k^{(m)}\}_k = \{V_1^{(m)}, \dots, V_{K_m}^{(m)}\}$. By using attribute values, an item $I_n \in \mathcal{I}$ can be corresponded to an attribute-value vector $\mathbf{V}_{I_n}^{(m)} = (v_{I_n,1}^{(m)}, \dots, v_{I_n,K_m}^{(m)})^\top$, where the value of $v_{I_n,k}^{(m)}$ is 1 when item I_n has attribute value $V_k^{(m)}$ and is 0 otherwise.

Let us denote users' attribute-of-focus of attribute type m at time t in a vector form as $\mathbf{f}_t^{(m)} = (f_{t,1}^{(m)}, \dots, f_{t,K_m}^{(m)})^\top$ $f_{t,k}^{(m)} \in \{0, 1\}$. If the user focuses on an attribute value $V_k^{(m)}$ at time t , $f_{t,k}^{(m)}$ is 1 and otherwise 0. Evaluation criteria are related to attribute values across several attribute types (e.g., "vegetable" and "green color"). However, the variations of combinations of attribute values become significant when the number of attribute types is large. To avoid the combinatorial complexity and keep the algorithm tractable, the proposed test is independently applied to each attribute type. That is, users' attribute-of-focus of each attribute type are first detected, and then the attribute-of-focus of all attribute types are obtained by merging those results. Therefore, in this section, we omit $^{(m)}$ that indicates attribute type m for simplicity.

We first model users' neutral browsing behavior by a multinomial distribution of observed attribute values. The multinomial parameter is given by a vector $\mathbf{p}$ for each attribute type. The k -th element of $\mathbf{p}$ denotes the probability that how likely the attribute value V_k is looked at under neutral browsing. Suppose we have gaze information as a sequence of 2-D points on displayed content. First, we obtain the sequence of gaze targets $\{i_t\}_t = \{i_1, \dots, i_T\}$ ($i_t \in \mathcal{I}$) from the sequence of gaze points. Note that, we define time t not by fixed-rate sampling used in eye-movement measurement but on the basis of the transition of gaze targets (i.e., $i_{t-1} \neq i_t$) because we specifically focus on users' comparison behavior. Then, we convert this sequence of gaze targets into a sequence of attribute values by referring to the attribute-value vector of each item i_t . As a result, we obtain a sequence of attribute-value vectors (*attribute-value sequence*) $\{\mathbf{V}_{i_t}\}_t = \{\mathbf{V}_{i_1}, \dots, \mathbf{V}_{i_T}\}$. Here, suppose an attribute-value sequence of length l are observed and a frequency distribution of observed attribute values in the sequence in the window is calculated. Let us denote the observed frequency in a vector form $\mathbf{x}_l$, whose elements are sum to l (we will describe following sections in detail). We simply refer to $\mathbf{x}_l$ as *gaze-attribute distribution*. Then, we assume that $\mathbf{x}_l$ in neutral browsing follows a probability distribution function $g(\mathbf{x}_l; l, \mathbf{p})$. That is, users' distinctive browsing behavior is identified by a p -value, which is the probability that the gaze-attribute distribution $\mathbf{x}_l$ or more biased distribution occurs in the window occurs under the assumption of neutral browsing. Note that a function g represents a distribution

of gaze-attribute distributions, and the small p -value means that the gaze behavior in the window is biased.

Let us consider two kinds of attributes: categorical attributes and ordinal attributes. In terms of the categorical attributes, a user may focus on each attribute value; for example, a user takes interests in “red” or “green” for a color. In contrast, regarding the ordinal attributes, a user may focus not on each attribute value but on a threshold of attribute values; for example, a user is concerned with a price “under \$15.” We first introduce the MSET for categorical attributes in Section 3.3.1 and then explain the MSET for ordinal attributes in Section 3.3.2 as an extension to the case of categorical attributes.

3.3 Multiscale Exact Test

3.3.1 In Case of Categorical Attributes

If an attribute type is categorical, each attribute value can be considered independently. In this case, distinctive browsing behavior is detected through the deviation from a multinomial distribution.

Modeling neutral browsing behavior Neutral browsing is defined as the users’ browsing behavior when they are not focused on any specific criterion. We simply assume that users browse catalog content uniformly when they are in neutral browsing. The model parameter $\mathbf{p} = (p_1, \dots, p_K)$ is calculated as follows, where p_k denotes how likely the attribute value V_k is to be looked at. Given a catalog content that contains N_k items having the k -th attribute value V_k , the k -th element of the multinomial parameter is derived as $p_k = N_k/N$.

MSET for categorical attributes We utilize the exact multinomial test to detect users’ distinctive browsing behavior for categorical attributes by given significance level for the test. To determine appropriate window sizes for the analysis, we apply the exact multinomial test to multiple time scales (Figure 3.2). Assume that we have an attribute-value sequence of length l and that the k -th attribute value exists x_k times in the sequence ($k = 1, \dots, K$). The probability of gaze-attribute distribution $\mathbf{x}_l = (x_1, \dots, x_K)$ of a time window in neutral browsing

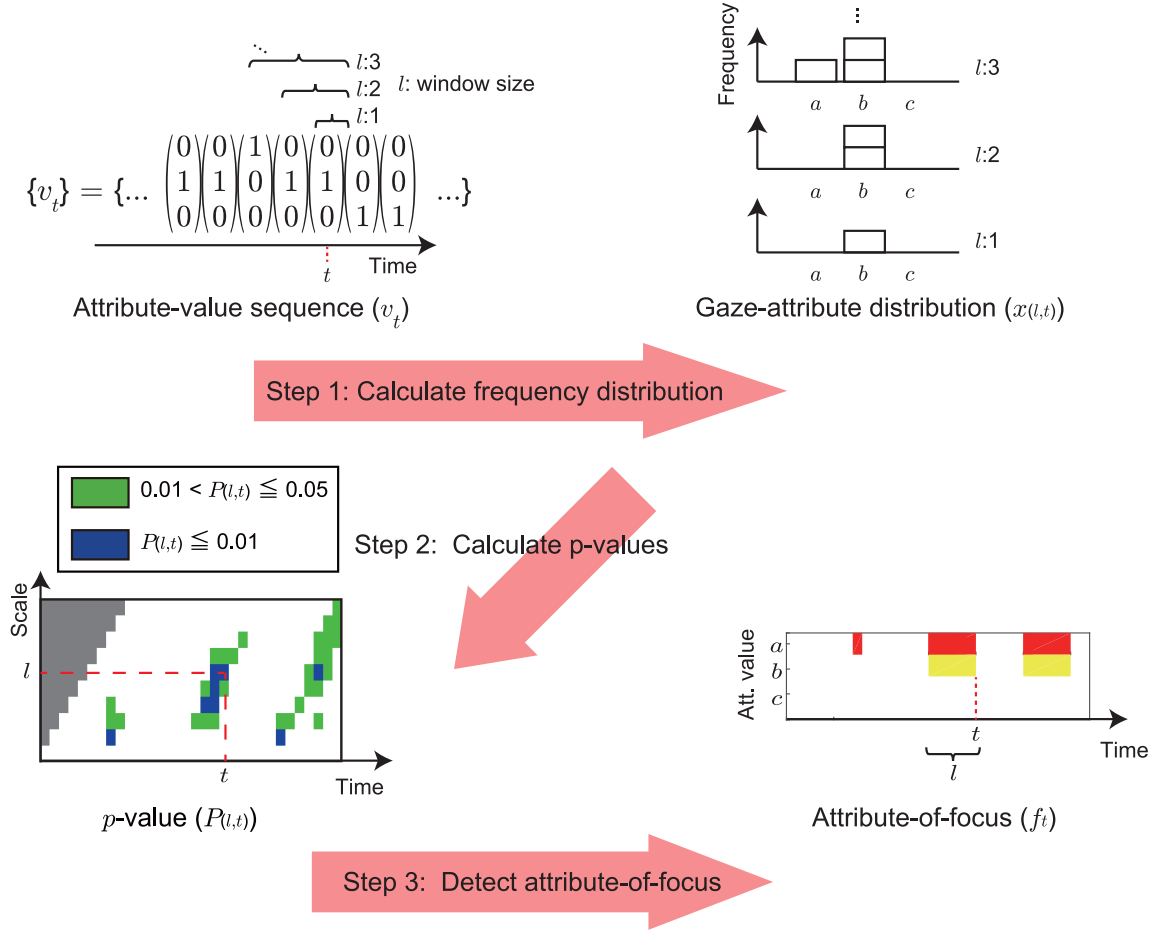


Figure 3.2: Proposed flow of the MSET to detect attribute-of-focus for categorical attributes. Different color in attribute-of-focus denotes different attribute value.

follows a multinomial distribution,

$$g_m(x_l; l, p) = l! \prod_{k=1}^K \frac{p_k^{x_k}}{x_k!}, \quad (3.1)$$

where p is the multinomial parameter defined in the previous paragraph.

Let $\{V_{i_{t'}}\}_{t'=t-l+1}^t$ be a subsequence of an attribute-value sequence $\{V_{i_t}\}_t$, where $1 \leq t' \leq T$, and l ($1 \leq l \leq 10$) is the window size. The gaze-attribute distribution $x_{(l,t)}$ is calculated for each subsequence (Figure 3.2, Step 1), where an additional subscript t is added to specify the position of the window. Then, distinctive windows can be identified by comparing the p -value to the significance level (Figure 3.2, Step 2). Here, the number of observation we assume is small (i.e., 10 or less); therefore, not Pearson's chi-squared test but the exact multinomial test

is applied. In contrast to the chi-squared test, which uses the approximation of the distribution assuming large sample size, the exact multinomial test directly calculates the p -value of observations using the original multinomial distribution:

$$P_{(l,t)} = \sum_{\hat{\mathbf{x}}: g_m(\hat{\mathbf{x}}; l, \mathbf{p}) \leq g_m(\mathbf{x}_{(l,t)}; l, \mathbf{p})} g_m(\hat{\mathbf{x}}; l, \mathbf{p}), \quad (3.2)$$

where the sum ranges over all outcomes as likely as, or less likely than, that observed. The MSET identifies “when” and “how long” the user’s gaze behavior is biased (i.e., time t and scale l) as shown in the left-bottom in Figure 3.2. This identification means that the users focus on some attribute values in $t - l \leq t' \leq t$. The MSET detects the users’ attribute-of-focus for each identified biased periods. Attribute-of-focus at time t , f_t , are determined by comparing a relative frequency of attribute values to the multinomial parameter. Specifically, if the relative frequency of the k -th attribute value x_k/l is higher than the multinomial parameter p_k , the attribute value V_k is regarded as a part of the set of attribute values of focus (i.e., $f_{t,k} = 1$) (Figure 3.2, Step 3).

Since the same t can be included in several windows (subsequences), multiple attribute-of-focus $f_{t,w}$ can be detected for time t , where the additional subscript w is an index of a window. Here, let $\mathcal{V}_{t,w}^f = \{V_k | f_{t,k} = 1\}$ be a set-form of attribute-of-focus $f_{t,w}$ obtained from window w . We consider $\mathcal{V}_t^f = \cup_w \mathcal{V}_{t,w}^f$ as the obtained attribute-of-focus at time t (we also use the corresponding vector-form f_t for the result).

3.3.2 In Case of Ordinal Attributes

In contrast to categorical attributes, people often concern a range of values for ordinal attributes (e.g., price is less than \$10). For this reason, it is better not to consider each attribute value independently but to divide the values into groups. To simply deal with ranges of values, we here assume that users have some threshold to divide attribute values into two groups and that the user focuses on one of those two groups. Under this assumption, we model users’ gaze behavior by binomial distribution and utilized the exact binomial test to detect users’ distinctive browsing behavior for ordinal attributes.

To determine a set of attribute values of focus, the binomial test is first applied to several thresholds. In this paper, all attribute values V_k ($k = 1, \dots, K - 1$) are chosen for possible thresholds.

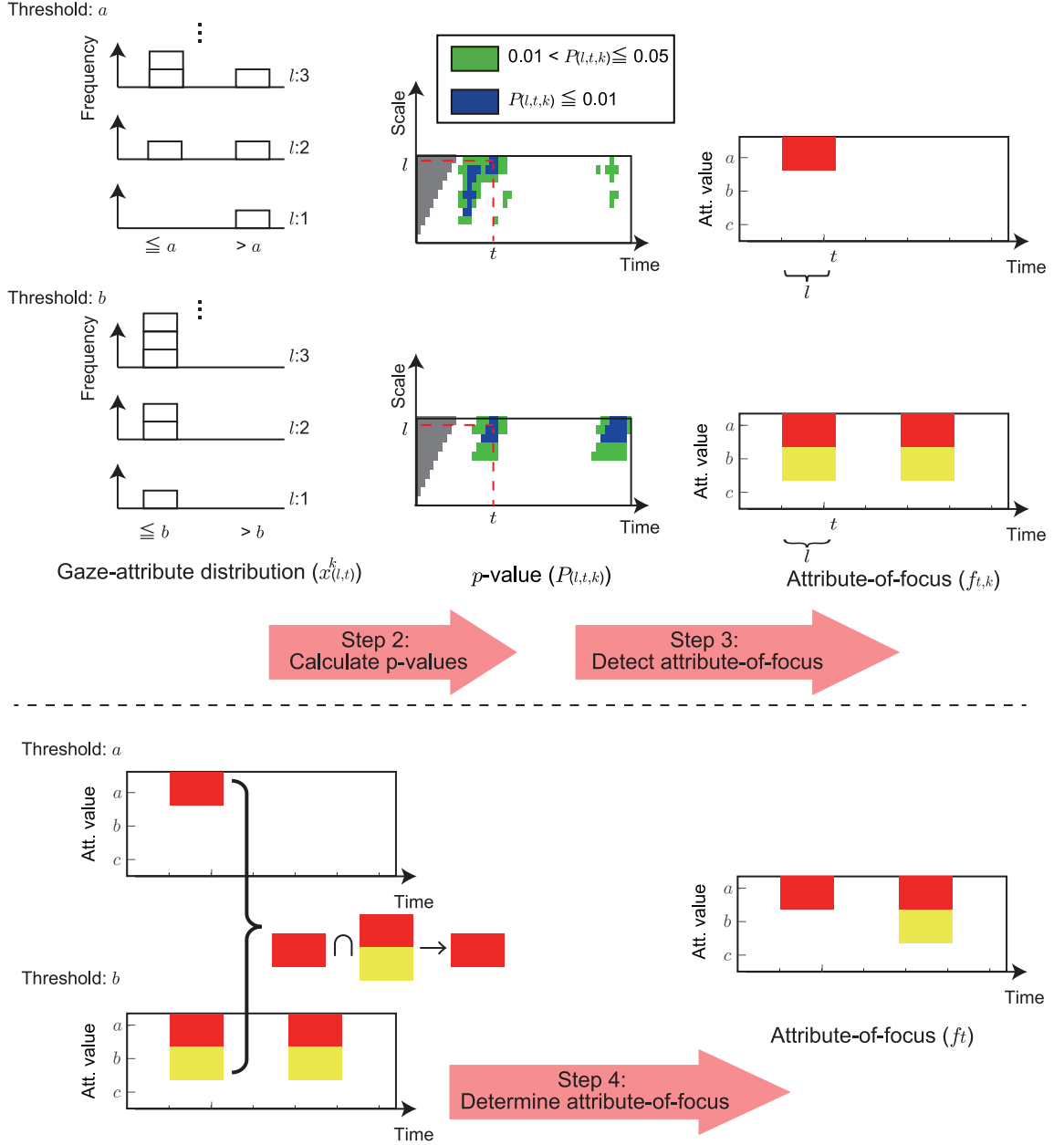


Figure 3.3: Proposed flow of the MSET to detect attribute-of-focus for ordinal attributes. Different color in attribute-of-focus denotes different attribute value.

Modeling neutral browsing behavior We prepare a binomial distribution for each threshold V_k with the model parameter $p_k = N_k/N$, where N_k is the number of items whose attribute value is higher than V_k . The p_k is the parameter describing how likely items having an attribute value higher than V_k are looked at under neutral browsing.

MSET for ordinal attributes In contrast to categorical attributes, we model users' gaze behavior by binomial distributions with different thresholds and utilize the exact binomial test for each of multiple window sizes. Assume we have an attribute-value sequence of length l , and the attribute values over threshold V_k exist x_k times in the sequence. The probability of gaze-attribute distribution $\mathbf{x}_l^k = (l - x_k, x_k)$ of a time window in neutral browsing follows a binomial distribution,

$$g_b(\mathbf{x}_l^k; l, p_k) = l! \frac{p_k^{x_k} (1 - p_k)^{l-x_k}}{x_k! (l - x_k)!}, \quad (3.3)$$

where p_k is a binomial parameter defined in the previous paragraph.

Similar to Section 3.3.1, we apply an exact test to each subsequence $\{\mathbf{V}_{t'}\}_{t'=t-l+1}^t$ of an attribute-value sequence. In the case of the ordinal attributes, p -value is calculated for each threshold V_k (Figure 3.3, Step 2) using

$$P_{(l,t,k)} = \sum_{\hat{\mathbf{x}}: g_b(\hat{\mathbf{x}}; l, p_k) \leq g_b(\mathbf{x}_{(l,t)}^k; l, p_k)} g_b(\hat{\mathbf{x}}; l, p_k). \quad (3.4)$$

The MSET also detects attribute-of-focus in each identified period by relative frequencies of attribute values similar to the categorical-attribute cases. If the relative frequency x_k/l is higher than the binomial parameter p_k , then the attribute values higher than V_k are regarded as attribute values of focus, i.e., $f_{t,k'} = 1$ ($k' > k$) (Figure 3.3, Step 3). On the other hand, if relative frequency distribution, $1 - x_k/l$, is higher than $1 - p_k$, the attribute values under or equal to the threshold V_k are regarded as attribute values of focus, i.e., $f_{t,k'} = 1$ ($k' \leq k$). Since the same t can be included in several windows, similar to the categorical-attribute cases, we take a union of multiple attribute-of-focus detected from all windows that include time t . Note that this union is obtained from each of thresholds V_k ($k = 1, \dots, K - 1$). Therefore, we denote the union as $\mathcal{V}_{t,k}^f$.

To merge the attribute-of-focus $\mathcal{V}_{t,k}^f$ ($k = 1, \dots, K - 1$) denoted with different thresholds, we take $\mathcal{V}_t^f = \cap_k \mathcal{V}_{t,k}^f$ so that the range of values becomes the narrowest (Figure 3.3, Step 4). This is because this test may detect a larger range of values as an attribute-of-focus (e.g., when the user focuses on only under \$5, the test detects not only under \$5 but also under \$10 as attribute values of focus).

3.4 Evaluation

We evaluated the proposed approach in terms of “When users are given a specific selection interests, how well attribute-of-focus are detected from the users’ gaze behavior correspond to the given tasks?” Two types of data were used for the evaluation: toy data and actual gaze data collected from eye tracking. Toy data was used to verify the concept of the proposed framework using ground truth data explicitly.

3.4.1 Toy Data

To prepare toy data, we simulated the following content and evaluation criteria. A content consisting of eight items was randomly generated for each session, where each item had three attribute types $\{X, Y, Z\}$ and each content satisfied the conditions described below. The proposed method was verified using two types of attributes: categorical and ordinal attributes.

Here, attribute type X was categorical attribute whose values were $\{a, b, c, d\}$, and attribute types Y and Z were ordinal attributes whose values were both $\{1, 2, 3, 4\}$. Two of eight items had the same attribute value for each of three types in a content (i.e., the number of items having a certain attribute value was equally set to two) so that the number of attribute values was not biased. With this content, four evaluation criteria C_q ($q = 1, 2, 3, 4$) were prepared. Each evaluation criterion C_q was given as a set of attribute values $\mathcal{V}_q^c$ such as “the value of attribute type X is a and the value of attribute type Y is higher than 2.” Here, the multinomial parameter of evaluation criterion p_q introduced in Chapter 2 was set uniformly for attribute values in evaluation criteria (i.e., $p_{q,k} = 1/|\mathcal{V}_q^c|$ if $V_k \in \mathcal{V}_q^c$, and 0 otherwise, where $|\cdot|$ is the cardinality of the set).

Two phases, a neutral browsing phase and a focused browsing phase, were prepared to verify whether the MSET could identify the focused phases and detect attribute-of-focus from those focused phases. Therefore, we simulated the situation where these two phases occurred alternately during users’ browsing behavior. The duration length of neutral browsing and focused browsing phases were set to 20 and 10, respectively, to verify the applicability of the MSET for short sequences.

In the neutral browsing phase, the probabilities of items being looked at were set to be uniform in accordance with the neutral-browsing model described in

Section 3.3. In contrast, users' browsing behavior in the focused browsing phases was generated based on the proposed model described in Chapter 2 as follows. The simulated user was first given a selection interest θ so that the user always focuses only one evaluation criterion as a task (i.e., $P(c_t = C_q) = 1$ for a specific evaluation criterion C_q and 0 otherwise). Then, attribute-of-focus at time t was given as a subset of an evaluation criterion $\mathcal{V}_q^c$ (i.e., $\mathcal{V}_t^f \subset \mathcal{V}_q^c$) in order to prepare a sequence of attribute-of-focus, $\{f_t\}_t$, for training data. Here, the number of attribute values $|\mathcal{V}_t^f|$ is as many as or slightly less than $|\mathcal{V}_q^c|$. For example, when the evaluation criterion was "the value of attribute type X is a and the value of attribute type Y is more than 2," attribute-of-focus at time t was given as "the value of attribute type X is a and the value of attribute type Y is 4." Under given attribute-of-focus, the probabilities of items being looked at were set according to the number of attribute values of items that match the attribute-of-focus. We generated five sessions for each evaluation criterion, i.e., we had $5 \text{ (sessions)} \times 4 \text{ (evaluation criteria)} = 20 \text{ sessions}$.

3.4.2 Results on Toy Data

An example of calculated p -values and the detected attribute-of-focus by the MSET on the toy data are shown in Figure 3.4 (a) and (b), respectively. Figure 3.4 (b) shows that the proposed multiscale detection method can detect distinctive periods with an attribute-of-focus in that period. The significance level was set to 1% to determine attribute-of-focus. In this example session, the specified evaluation criterion was "the value of attribute type X is b, and the value of attribute type Z is more than 2." These results show that value b of attribute X and values 3, 4 of attribute Z were successfully detected in focused browsing periods (gray periods in Figure 3.4). However, the MSET did not fully detect all attribute values specified by evaluation criterion. This was because the simulated user did not always focus on all attribute values in an evaluation criterion, but focused on a part of the evaluation criterion as an attribute-of-focus (i.e., simulated user sometimes did not focus on value 3 of Z).

To confirm the accuracy of the detection results of participants' comparison behavior by MSET more quantitatively, similarities between detected attribute-of-focus and task related attributes are shown in Figure 3.5. In these results, the similarities were defined by cosine similarities:

$$\text{Sim}(f, \hat{f}) = \cos(f, \hat{f}), \quad (3.5)$$

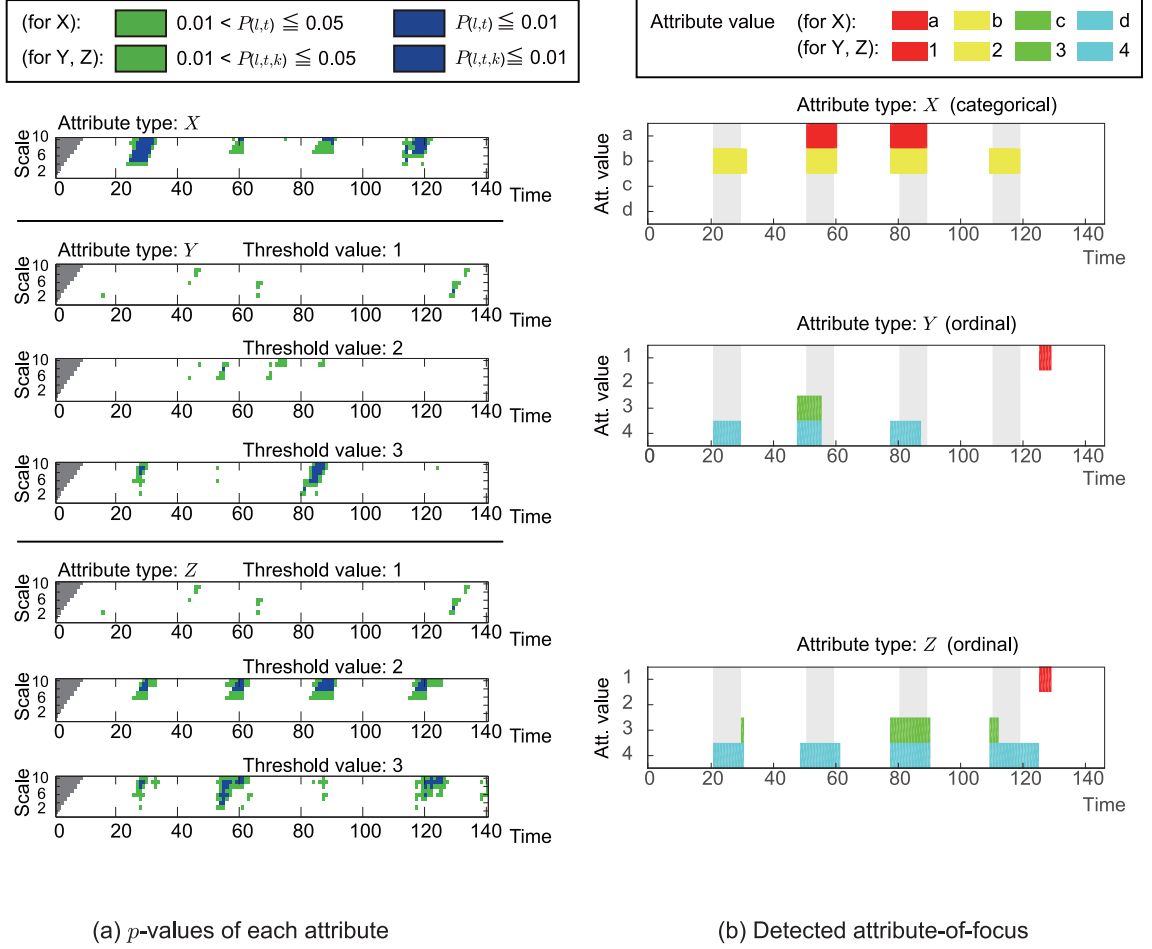


Figure 3.4: An example of the results of the MSET on toy data: (a) p -values and (b) identified distinctive browsing period and detected attribute-of-focus. Each figure of (a) shows when and how long distinctive browsing occurs. Since attribute type Y and Z are ordinal attributes, three figures of p -values are shown for three thresholds. Each figure of (b) corresponds to each attribute type, and different colors indicate different attribute values. White and gray periods show neutral browsing phase and focused browsing phases, respectively.

3.4. EVALUATION

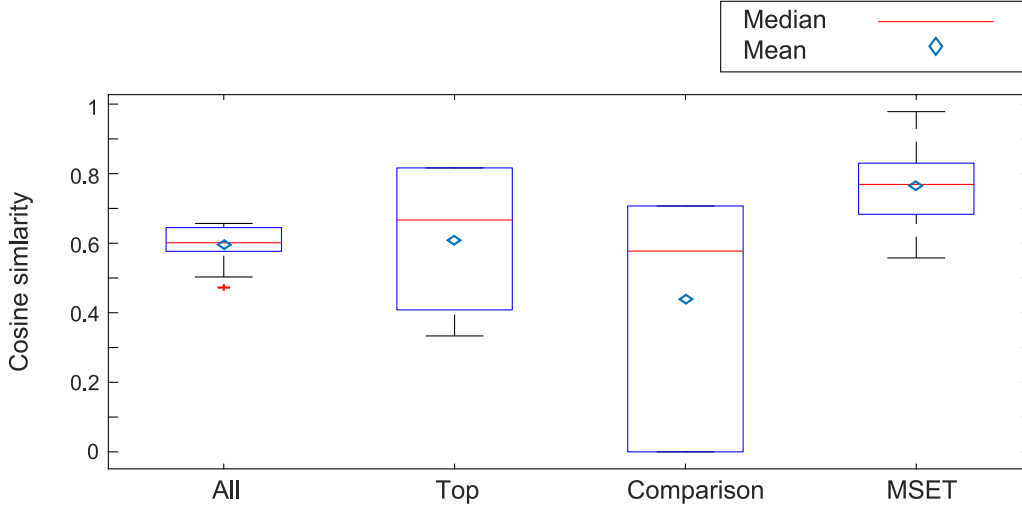


Figure 3.5: Similarities between detected attribute of focus and task related attribute values

where f denotes a detected attribute-of-focus and $\hat{f}$ denotes a vector-form of task-related attribute values of the given task. Here, all detected attribute-of-focus were regarded as the attribute values on which the participant focuses, that is, $f_k = \sum_t f_{k,t}$. The purpose of the MSET is to detect users' comparison behavior, three baseline method were prepared: *All*, *Top*, *Comparison*.

All attribute values of focus are determined by the gaze-attribute distributions of whole sequence, $f_k = \sum_t v_{i_t,k}$.

Top attribute values of focus are determined by the attribute values of the most browsed items by the participant, $f = V_{I_{top}}$.

Comparison attribute values of focus are determined by the attribute values that commonly held by the most browsed and the second most browsed items by the participant, $f_k = f_{k,I_{top}} f_{k,I_{top2}}$.

The difference between the MSET and these baseline methods were enough significant ($p < .01$). These results show that by the MSET successfully detects users' comparison behavior with its attribute values of focus.

3.4.3 Actual Gaze

We conducted an experiment to evaluate the proposed method with actual gaze data obtained from the cooperation of 37 participants (18 males and 19 females,

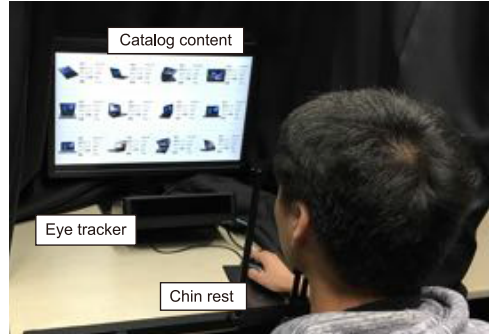


Figure 3.6: Experimental environment

university students, ranging from age 19 to 34, mean and standard deviation of age were 22.3 and 2.9, respectively).

Experimental Design

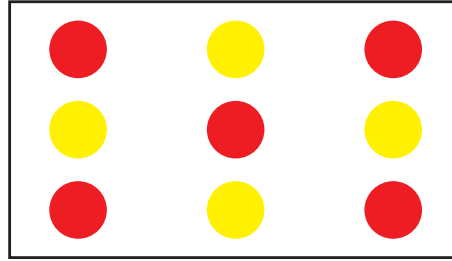
In the experiments, participants were asked to select one item (PC, laptop computer) from 12 displayed items on a screen (see Figure 3.6), and an eye tracker¹ under the display was used to measure participants' eye movements. Each item had five attribute types: *price*, *screen size*, *CPU score*, *memory capacity*, and *weight*. All attribute types were ordinal, and each attribute type could take three attribute values, respectively. A content for each session was prepared so that four out of 12 items had the same attribute value for each of five types in each content. We gave participants tasks that specified requirements of "situations" and "purpose" for each session; for example, "Please assume that you will use your primary PC at home to watch movies or to play games. Which PC do you think is best for that situation?" By giving tasks, we assumed that the participants' selection interests were constant during decision making. Common three tasks were prepared and used for all participant. Each participant conducted three sessions corresponding to the three tasks, where the order of the tasks was randomized. Therefore, these three tasks were considered to be the evaluation criteria. These requirements were not explicit constraints but were implicitly related to several attribute values. For instance, a PC for playing games or watching movies required "a high CPU score and large memory capacity." We refer to these values as *task-related attribute values*. Table 3.1 summarizes three requirements given as tasks and task-

¹Tobii X120 Eye Tracker: Freedom of head movement is $300 \times 220 \times 300$ mm, sampling rate is 60 Hz, and accuracy is 0.5 degrees.

3.4. EVALUATION

Table 3.1: Given tasks and attribute values expected to be related to the tasks

Given situation	Purpose of selection	Task-related attribute values
To use at home	To play games	High CPU score $V_3^{(CPU)}$ Large memory capacity $V_3^{(mem)}$
To carry outside	To take a note	Small screen size $V_1^{(siz)}$ Right weight $V_1^{(wei)}$
To buy with low cost	To see web pages	High CPU score $V_3^{(CPU)}$ Low price $V_1^{(pri)}$



Red Points: for calibration

Yellow Points: for verification

Figure 3.7: Points used for the calibration of the eye tracker. This figure express only where these points are located, and size is exaggerated.

related attribute values of each of the tasks. Therefore, letting $\mathcal{V}_q^c$ ($r = 1, 2, 3$) be the set of task-related attribute values of task r , we considered $\mathcal{V}_q^c$ as the grand truth of the aspect learning. To elicit participants' comparison behavior, we set the number of items that fulfill each evaluation criterion $\mathcal{V}_q^c$ in each session as two so that participants could not uniquely decide on one item on the basis of a given task.

To decrease the influence of the spatial layout, we randomized the position in the content. In this experiment, the number of sessions we obtained were 37 (participants) $\times$ 3 (tasks) = 111 (total sessions).

Procedure

Each participant was first asked to sit facing the display and to position their face to be aligned with the chin rest (Figure 3.6). The eye tracker was then calibrated for each participant. After the calibration, we explained the content and proce-

dures of this experiment to participants.

The experimental procedure of each session consisted of the followings four steps.

Step 1 Participants were explained a task.

Step 2 The accuracy of eye tracker's calibrated parameter was confirmed.

Step 3 Contents were displayed, and the participants were asked to choose items.

Step 4 After participants chose an item, we asked them which PC they had chosen.

Calibration before Step 1, and calibration conformation in Step 2, as shown in Figure 3.7, five points (depicted by red points) on the display were used to calibrate the eye tracker. Then, the other four points (depicted yellow points in Figure 3.7) on the display were used to measure calibration error. If this error was over a threshold, we retried the calibration until the error was under the threshold.

The purpose of this study is modeling gaze behavior while user compares and explores displayed items. However, when we human browse a content and make a choice from it, most of us glance at it and try to grasp what items exist and where they are located as discussed in Section 1.2.5. Hence, we try to decrease the effect of this phase, what we refer to as the glimpse phase, so that we can observe eye gaze patterns more related to the comparison and exploration of displayed items. For instance, Hirayama et al. [Hirayama et al., 2010] took an approach that all articles are displayed at few seconds intervals in turns before measurement of gaze data for analysis. Therefore, in Step 3, to explicitly distinguish users' glimpse phase from comparison behavior in decision making [Russo and Leclerc, 1994, Shi et al., 2013], we followed their approach. That is, all items were displayed at intervals of three seconds in turn before the measurement of gaze data for analysis. Since we did not limit decision time, participants browsed content and compared items before deciding on one item.

3.4.4 Results on Gaze Data

We first investigate the basic properties of the decision-making sessions such as the duration length and summarize some statistics found from the results of the MSET to better understand the browsing behavior seen in the experiments.

3.4. EVALUATION

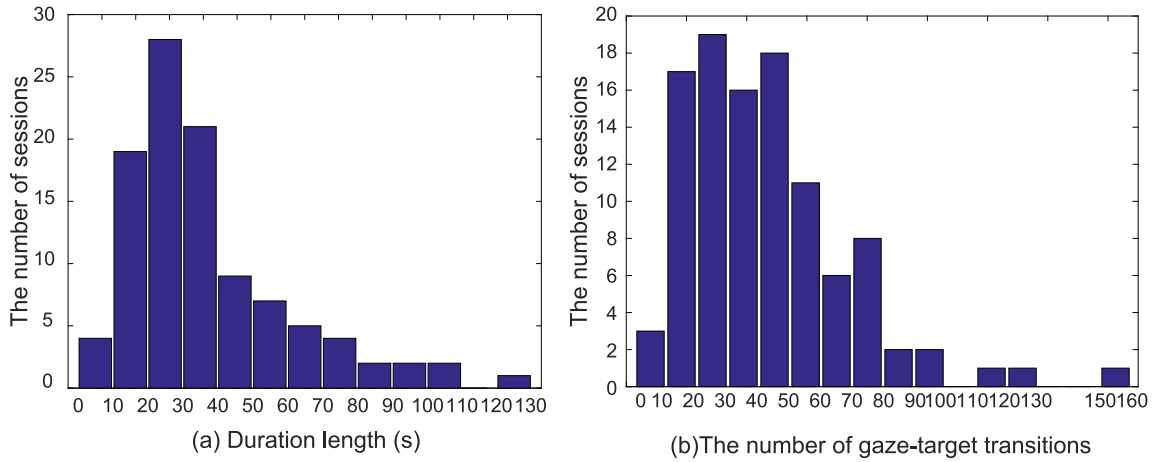


Figure 3.8: (a) Histogram of session duration length. The mean of duration length was 33.3 seconds and the median was 25.6 seconds. (b) Histogram of the number of gaze target transitions in each session. The mean number of times was 37.8, and the median was 33.

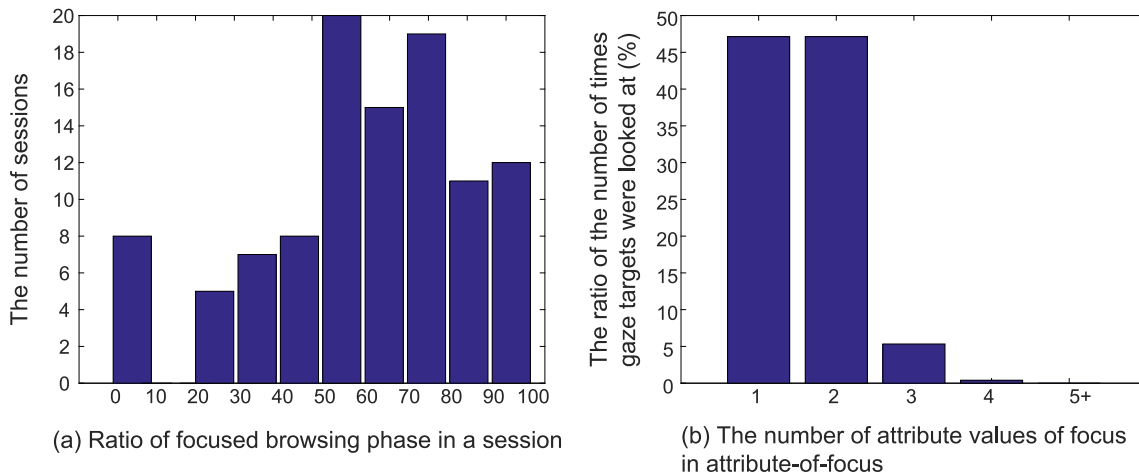


Figure 3.9: (a) Histogram of the ratio of focused browsing phase in each session detected by the MSET. The mean of the ratio was 64.9%, and the median was 69.4%. Note that these durations were not actual sampling times but were based on the number of times targets were looked at in a session. (b) Histogram of the number of attribute values of focus in attribute-of-focus detected by the MSET for all session. The mean number of focused value was 1.59, and the median was 2.

Figure 3.8 (a) and (b) shows the histograms of session duration length and the number of gaze-target transitions in each session, respectively. From these figures, we can see that the participants' comparison in several sessions were short. One of possible reasons is that they did not need a time for comparison due to the procedure of the experiments. That is, because we gave a specific task for each

session and explicitly distinguished screening phase, the participants might have almost decided items during screening phase.

Figure 3.9 (a) shows the ratio of the total length of focused browsing phases detected by the MSET. Note that the length was based on the number of times gaze-target transitions. This result indicates that participants did not always compare items and sometimes only examines several items to acquire information of the displayed content. In such situations, the participants' browsing behavior might not reflect any evaluation criteria. Figure 3.9 (b) shows the histogram of the number of detected attribute values by the MSET. We can see that participants focused on one or two of 15 attribute values while browsing multi-attribute content. Those results show that participants' attribute-of-focus could be observed temporarily (Figure 3.9 (a)) and attribute-of-focus is partial of attribute values of items (Figure 3.9 (b)).

An example result of calculated p -values and the detected attribute-of-focus from actual gaze data is shown in Figure 3.10. The significance level was set to 1%, similar to Section 3.4.2. The task was to select an item with higher CPU score, $V_3^{(CPU)}$, and larger memory capacity, $V_3^{(mem)}$. Here, the subscripts 3, denote the largest attribute values in each of types, i.e., CPU scores and memory capacity. In this session, we can see that the participant first focused on high CPU score and then compared items with high CPU score and large memory capacity.

To confirm the accuracy of the detection results of participants' comparison behavior by MSET more quantitatively, similarities between detected attribute-of-focus and task related attributes are calculated same as toy data (see Figure 3.11). In these results, the similarities were calculated by cosine similarity by Equation (3.5). The difference between the MSET and these baseline methods were significant (p -values were 1.4×10^{-2} , 2.9×10^{-7} , 1.9×10^{-7} , respectively). These results show the MSET successfully detects participants' attribute-of-focus.

3.5 Discussion

In this section, we discuss the limitations of the proposed framework.

Detection by the significance level The MSET identifies users' comparison behavior by comparing p -value and the significance level. However, this identification ignores the degree of the bias of the users' gaze behavior, and therefore the identification results depend on the significance level we sat. To address this is-

3.5. DISCUSSION

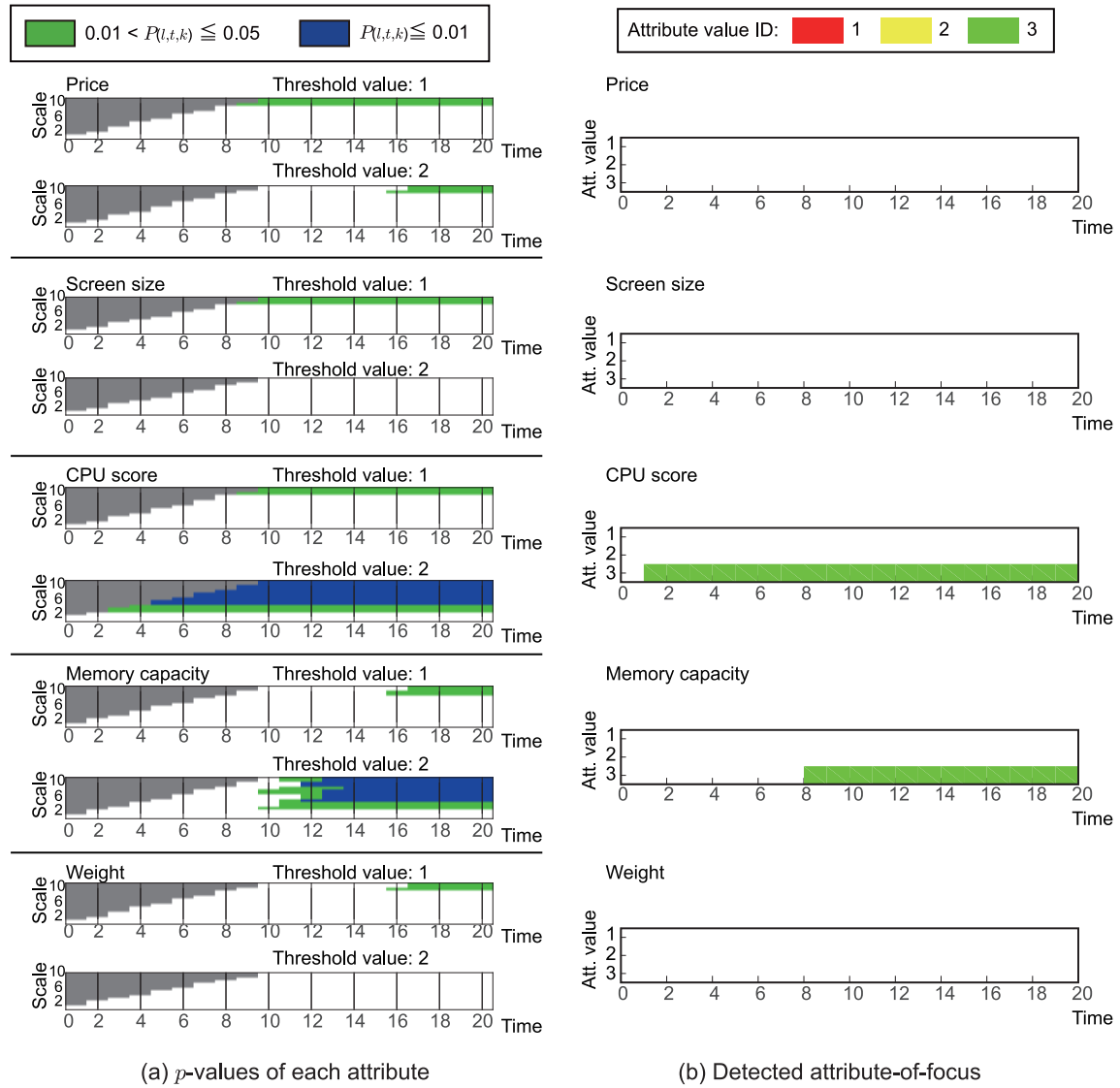


Figure 3.10: An example of the results of the MSET on gaze data: (a) p -values and (b) identified distinctive browsing period and detected attribute-of-focus. Each figure of (a) shows when and how long distinctive browsing occurred. Since all attributes are ordinal, two figures of p -values for two thresholds. Each figure of (b) corresponds to each attribute type.

sue, possible approach may be to utilize p -value itself as a probability of the bias and to extend the bias detection approach to an estimation of probability distribution of attribute-of-focus.

Focus on a range of ordinal attribute values In the analysis of the MSET, ordinal attribute values are divided into two groups with one threshold, but this is not

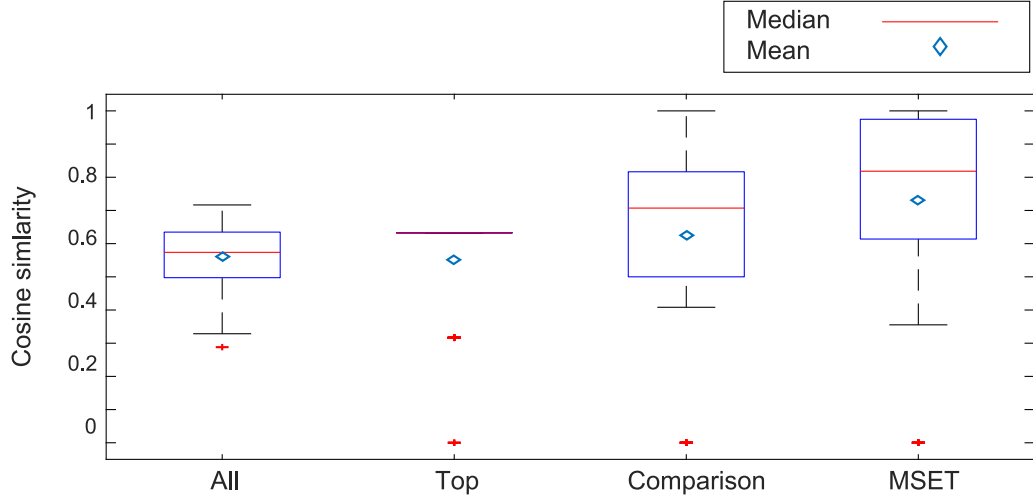


Figure 3.11: Similarities between detected attribute of focus and task related attribute values from actual data

appropriate when a user focuses on a middle range of attribute values. Although our approach can detect a middle range of users' focus by taking an intersection of detection results (e.g., "more than \$5" and "less than \$20"), our approach is not designed to directly detect a middle range of attribute values as attribute-of-focus. One simple approach to overcome this limitation may be introducing two thresholds to divide attribute values. However, modeling all range of attribute values becomes a combinatorial problem.

The effect of spatial layout While we simply assume that users uniformly browse displayed items on a screen in the neutral browsing, the position of each item in the content may affect users' gaze behavior (e.g., center bias [Borji, 2012]). Therefore, taking account of the effect of spatial layout to neutral browsing model is expected to increase the accuracy of the attribute-of-focus detection.

Chapter 4

Gaze Behavior Model with Attribute-of-Focus for Aspect Learning

4.1 Introduction

As mentioned in Chapter 2, constructing the state space of interests is one important issue to realize interactive assistance systems because effectiveness of interactive assistance depends on the state space. In this chapter, we propose an approach to obtain possible evaluation criteria from users' comparison behavior because users' comparison behavior reflects the users' evaluation criteria. To address the issues that how can we obtain evaluation criteria by taking users' comparison behavior into account, we propose a framework for obtaining evaluation criteria from gaze behavior by taking account of users' *focus* on attribute values.

The proposed framework consists of two steps as shown in Figure 4.1. The first step highlights short-term biased segments during users' content browsing using the MSET introduced in Chapter 3. Then, the results is used in the second step with the *Aspect-Oriented Gaze-Behavior Model*. In this chapter, we describe how to adopt gaze behavior model described in Chapter 2 to the MSET.

In our framework, the MSET first detects users' attribute-of-focus for each attribute type from users' gaze behavior (the left side of Figure 4.1). To obtain evaluation criteria after detecting users' attribute-of-focus, we apply the aspect-oriented gaze-behavior model (the right side of Figure 4.1). Note that the proposed model has a hierarchical structure similar to the AHP (shown in Figure 2.3),

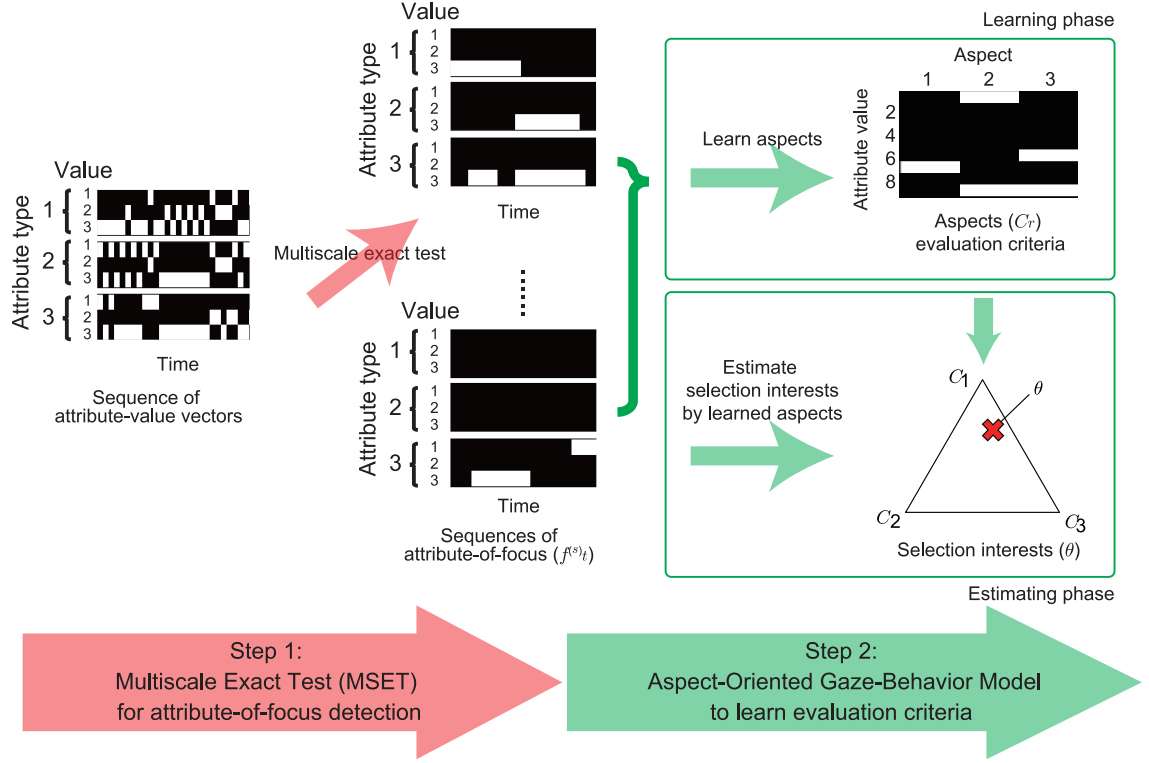


Figure 4.1: Proposed flow of two-step approach to obtain evaluation criteria from users' gaze behavior. The first step detects users' attribute-of-focus, and the second step learns evaluation criteria from detected attribute-of-foci.

where each aspect is described by the degree of association with each of attribute values.

The contribution of this chapter is as follows: We propose a novel framework of learning evaluation criteria in a bottom-up fashion by introducing a detection method of attribute-of-focus in order to take into account users' dynamic focus change on attribute types and attribute values.

As described in Section 2.7.2, by using aspects learned from a large amount of users' gaze data of comparison behavior, a user's selection interests can be estimated from the user's newly observed gaze data through probabilistic inference.

4.2 Approach for Learning Evaluation Criteria

With the representation of aspects, the goal of this chapter is to learn the degree of association between aspects (evaluation criteria) and attribute values from users' attribute-of-focus. This learning phase is the second step of our framework (see

4.2. APPROACH FOR LEARNING EVALUATION CRITERIA

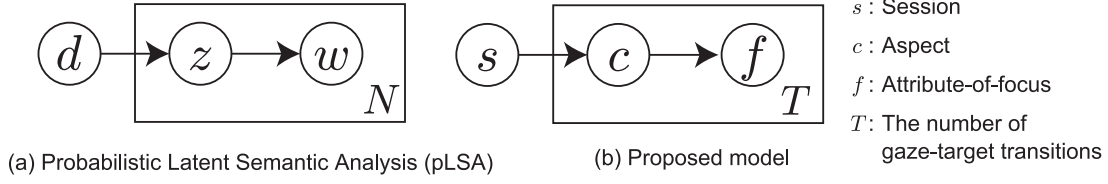


Figure 4.2: Graphical models of pLSA and proposed model

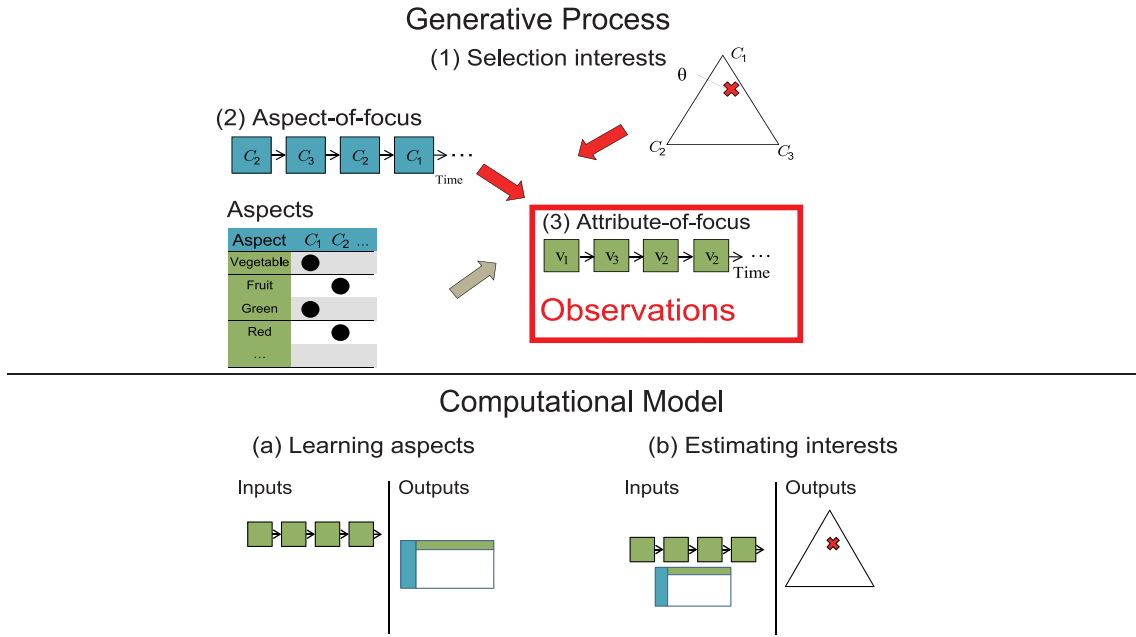


Figure 4.3: A probabilistic generative process of attribute-of-focus from users' selection interests

also Figure 4.1).

We apply topic models to gaze behavior in order to learn evaluation criteria as a form of aspect and propose an aspect-oriented gaze-behavior model in Chapter 2. Figure 4.3 shows the overview of the proposed model. Its top side shows the generative process of attribute-of-focus from selection interests. In our model, the sequence of attribute-of-focus $\{f_t\}_t$ is considered as a users' gaze behavior, that is, the observation $\mathcal{O}$ in Chapter 2 is the sequence of attribute-of-focus as described in Chapter 3. Corresponding to a document, we define a *session* that consists of a sequence of attribute-of-focus detected by the MSET during one decision making. Besides, by corresponding aspects of items to topics in the pLSA, the proposed gaze-behavior model has the similar structure of pLSA as shown in

Figure 4.2.

In contrast to the pLSA, which considers each word as an observation, the observation of our model at each time step is “attribute-of-focus” from user’s internal selection interest. Therefore, its observation model should be able to represent joint distribution of attribute values constituting attribute-of-focus. One possible approach for modeling such a multivariate joint probability is the use of a naive Bayes model consisting of multiple categorical distributions, where each of the distributions corresponds to each attribute type and expresses how likely each value of the type becomes attribute-of-focus. However, the use of a naive Bayes model is not appropriate here because, in many situations, users simultaneously focus on not all attribute types but only a part of them due to “attention resource” as discussed in Section 4.1. On the other hand, multiple attribute values can be detected as attribute-of-focus of a single attribute type using the MSET described in Chapter 3, and we want the model to express this multiplicity of focus as well.

To take both the partiality and multiplicity of users’ focus into account, we explicitly model users’ “attention resource” in our gaze-behavior model. Specifically, we first consider a multinomial distribution on “all the attribute values” as the observation model and introduce the number of attribute values of simultaneous focus, n_t , as a parameter of the attention resource at time t . Assuming the parameter n_t is given from the result of the MSET and used as the number of “trial of focus” at a time, we successfully represent users’ focus on partial attribute types and multiple attribute values. Note that the same attribute value does not appear more than once in our observation of attribute-of-focus, while the multinomial distribution also allows multiple observation of the same value up to n_t . In spite of that, the approximation with multinomial distribution is still reasonable in our case since n_t is much smaller compared to the total number of attribute values (K is used for notation in the next subsection).

With this model, suppose the sequences of attribute-of-focus are observed, aspects can be learned by maximum likelihood parameter estimation of the probabilistic model. Once aspects are learned, users’ selection interests can also be estimated (the right side of Figure 4.3).

4.3 Formulations of the Content-Browsing Situation

We here formulate users’ internal states during content browsing by introducing additional notations to those used in Chapter 2 and Chapter 3. We redefine the

attribute-of-focus as $\mathbf{f}_t = (f_{t,1}, \dots, f_{t,K})^\top$ and model the aspect as a multinomial parameter vector, whose number of elements is K , for the convenience. Besides, by using the number of attribute values of focus n_t at time t , introduced in the previous section, we assume that attribute values of focus, which is represented as an attribute-of-focus $\mathbf{f}_t$, is obey a multinomial probability distribution $h(\mathbf{f}_t; n, \mathbf{p})$, conditioned by n_t and a multinomial parameter of an aspect $\mathbf{p}$.

4.4 Generative Process of Attribute-of-Focus

The generative process of gaze behavior is modeled as follows (Figure 4.3). During session s , a user is assumed to have a selection interest modeled by $\theta(s)$ and first focuses on an *aspect-of-focus* at each time t based on users' selection interest ((1) to (2) in Figure 4.3) similar to Chapter 2.

To take attribute-of-focus into consideration, the latter part of aspect model is extended from Chapter 2. We assume that the user turns attention to attribute values, represented as an attribute-of-focus $\mathbf{f}_t$, conditioned by the aspect-of-focus c_t . When c_t is determined to be C_q , $\mathbf{f}_t$ is assumed to obey the multinomial probability distribution $P(\mathbf{f}_t | c_t = C_q) = h(\mathbf{f}_t; n_t, \mathbf{p}_q)$ ((2) to (3) in Figure 4.3). Note that the function h is also conditioned by n_t , the number of the attribute values of focus. In our framework, n_t is assumed to be given as the estimated number of attribute-of-focus by the MSET. $h(\mathbf{f}_t; n_t, \mathbf{p}_q)$ is derived as

$$h(\mathbf{f}_t; n_t, \mathbf{p}_q) = n_t! \prod_{k=1}^K \frac{p_{q,k}^{f_{t,k}}}{f_{t,k}!} = n_t! \prod_{k=1}^K p_{q,k}^{f_{t,k}}, \quad (4.1)$$

where $n_t = \sum_k f_{t,k}$. We assume that the user generates the sequence of attribute-of-focus by repeating these generative process for every time step in a session with constant selection interests $\theta(s)$.

The joint probability of an aspect-of-focus and attribute-of-focus in session s is derived as follows by assuming the conditional independence of $\mathbf{f}_t$ and s given c_t :

$$\begin{aligned} P(\mathbf{f}_t, c_t = C_q | s) &= P(\mathbf{f}_t | c_t = C_q) P(c_t = C_q | s) \\ &= h(\mathbf{f}_t; n_t, \mathbf{p}_q) \theta_q(s). \end{aligned} \quad (4.2)$$

Hence, the probability that an attribute-of-focus $\mathbf{f}_t$ is observed in session s is given

by

$$P(\mathbf{f}_t|s) = \sum_q h(\mathbf{f}_t; n_t, \mathbf{p}_q) \theta_q(s). \quad (4.3)$$

4.5 Estimation of Model Parameters

In this section, we list the parameters to be estimated. When we have S sessions of decision making, parameters to be estimated are $\theta(s)$, users' selection interests in each session s , and $\{\mathbf{p}_q\}_q$, multinomial parameters that characterize aspects. Here, the probability of attribute-of-focus $\mathbf{f}_t$ in session s is given by Equation (4.3). Remind that the parameter n_t is given as described in Section 4.4. Given the parameters $\theta(s)$ and $\{\mathbf{p}_q\}_q$, the likelihood of a sequence of attribute-of-focus $\{\mathbf{f}_t^{(s)}\}_t$ in session s is computed as:

$$\begin{aligned} L(\{\mathbf{f}_t^{(s)}\}_t) &= \prod_t P(\mathbf{f}_t^{(s)}|s) \\ &= \prod_t \sum_q h(\mathbf{f}_t^{(s)}; n_t, \mathbf{p}_q) \theta_q(s). \end{aligned} \quad (4.4)$$

That is, optimization equation is described as follows:

$$\begin{aligned} &\underset{\{\theta(s)\}_s, \{\mathbf{p}_q\}_q}{\text{maximize}} && \sum_s \sum_t \log \sum_q h(\mathbf{f}_t; n_t, \mathbf{p}_q) \theta_q(s) \\ &\text{subject to} && \sum_q \theta_q(s) = 1 \quad \forall s \\ &&& \sum_k p_{q,k} = 1 \quad \forall q. \end{aligned} \quad (4.5)$$

4.6 Evaluation

We evaluated the proposed framework in terms of two points: (1) When users are given a specific selection interests, how well evaluation criteria obtained from the users' gaze behavior correspond to the given selection interests; and (2) how well selection interests are estimated from gaze behavior? The data for evaluation were same as Chapter 3.

4.6. EVALUATION

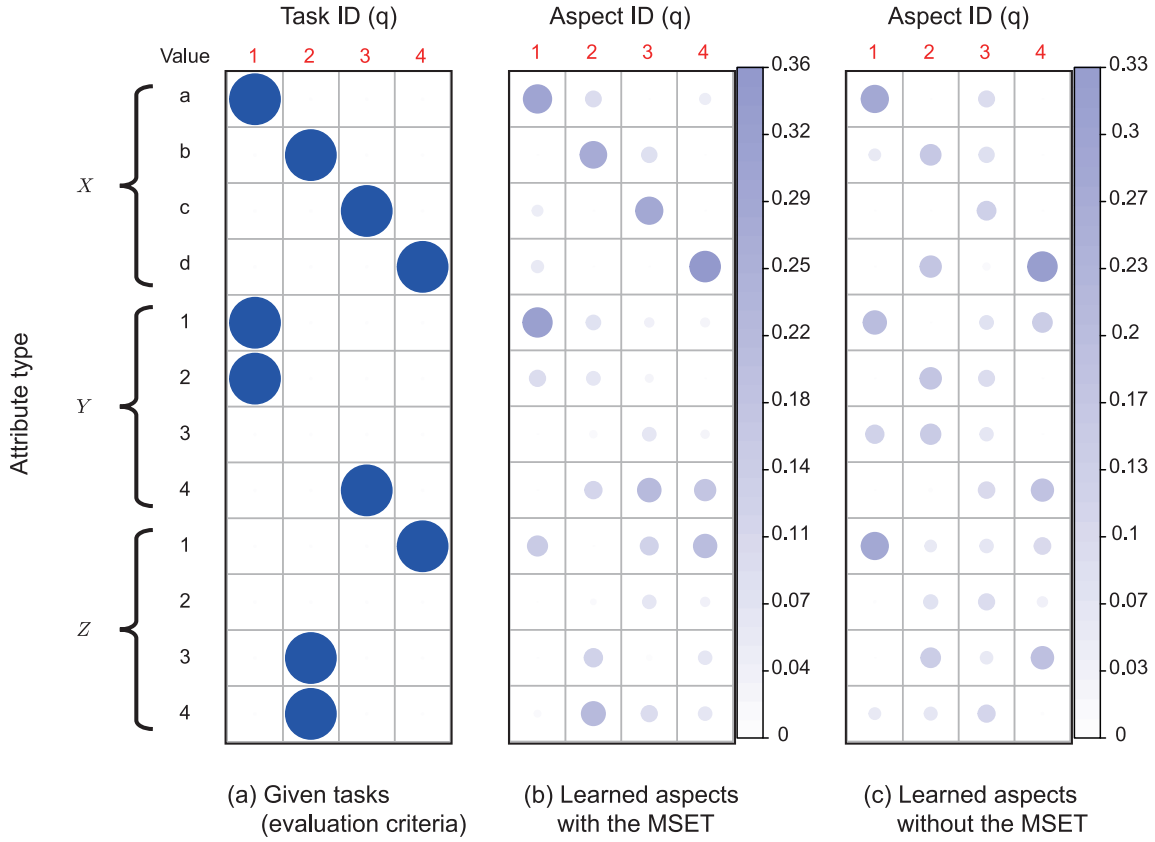


Figure 4.4: Given tasks and learned aspects with and without the MSET on toy data. The size and color of each dot depict the value of the multinomial parameter $p_{q,k}$, which corresponds to the degree the aspect C_q is related to attribute values V_k .

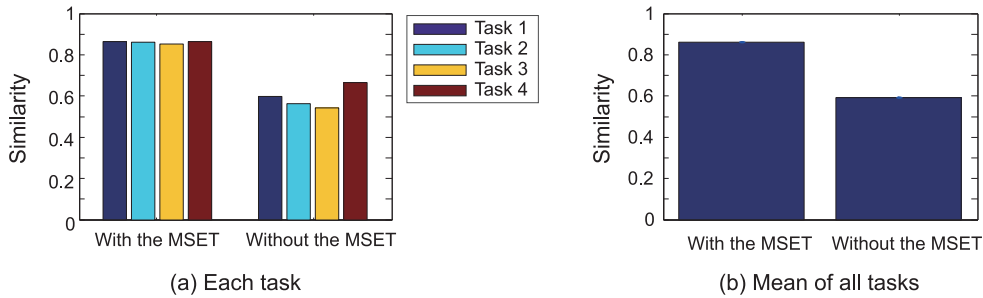


Figure 4.5: Similarities between given evaluation criteria and learned aspects with and without the MSET. Each bar of the (a) shows the similarity between learned aspect and the evaluation criterion given as the task. (b) shows the mean of similarities of all tasks.

Table 4.1: Results of evaluation criterion estimation on toy data. Number in each cell shows the number of sessions that were estimated as engaged task q ($q = 1, \dots, 4$). “No comparison” means the MSET could not detect a bias of gaze behavior in the session.

		Estimated task				
		1	2	3	4	No comparison
Actual task	1	5	0	0	0	0
	2	0	5	0	0	0
	3	0	1	4	0	0
	4	0	0	0	5	0

4.6.1 Toy Data

Figure 4.4 (a) shows attribute values specified as given tasks (i.e., evaluation criteria) and Figure 4.4 (b) and (c) show the learned aspects by the parameter estimation introduced in Chapter 2, with and without the use of the MSET. Here, for the “without the MSET,” not the attribute-of-focus (e.g., Figure 3.1 (b)) but the original attribute-value sequence obtained from gaze data (e.g., Figure 3.1 (a)) was directly used for the input of the aspect learning. This result shows that the learned aspects with the MSET had much clear association with the attribute values that correspond to one of given tasks. To confirm the effect of the MSET more quantitatively, similarities between learned aspects and evaluation criteria were calculated as shown in Figure 4.5. In these results, the similarities were defined by the cosine similarity of two parameter vectors,

$$Sim(q, \hat{q}) = \cos(\mathbf{p}_q, \hat{\mathbf{p}}_{\hat{q}}), \quad (4.6)$$

where $\mathbf{p}_q$ denotes a multinomial parameter of a learned aspect q ($q = 1, 2, 3, 4$) and $\hat{\mathbf{p}}_{\hat{q}}$ denotes the evaluation criterion of given task $\hat{q}$ ($\hat{q} = 1, 2, 3, 4$). Here, the pair-matching was determined so that the sum of the similarities of four pairs became the largest among all possible pairs $(q, \hat{q})$. These results show that the MSET much improves the distinctiveness of learned aspects.

Once aspects are learned, the proposed framework can also estimate users’ selection interests as shown in Figure 4.3 (b). Table 4.1 shows the results of selection-interests estimation. In this experiment, we assumed that users’ selection interests were related to only one evaluation criterion. Therefore, the estimation can be considered as the task classification in which the result is given by the maxi-

imum probability of estimated selection interests, $\arg \max_q P(C_q|s)$. Since the proposed method first detects attribute-of-focus by the MSET and estimates selection interests from detected attribute-of-focus, this approach can estimate selection interests only when biased gaze behavior is detected. “No comparison” means that the MSET could not detect any window with biased gaze behavior of gaze behavior. In these simulation results, the MSET successfully detected comparison behavior as a bias of gaze behavior for all sessions, and the accuracy of evaluation criterion estimation was 95%. This result shows that the proposed approach can also estimate users’ selection interests.

4.6.2 Actual Gaze Data

The results of the learned aspects with and without the MSET are shown in Figure 4.6 (b) and (c), respectively, and can be compared qualitatively with the task-related attribute values shown in Figure 4.6 (a) in terms of the degree of association with the attribute values. In the aspect learning, the number of aspects was set to be the same as that of tasks so that aspects corresponding to the given tasks were obtained. These results show that the learned aspects were more distinct when the MSET was applied, and the attribute values highly associated with each aspect seem to be similar enough to the task-related attribute values for all the three tasks.

To confirm the effect of the MSET quantitatively, similarities between the learned aspects and the given tasks were calculated in the same manner as the evaluation with toy data (see Section 4.6.1 and Equation (4.6)). Figure 4.7 shows the similarity scores between each task and the corresponding aspects found by the matching of the given tasks and the learned aspects similar to Section 4.6.1. Compared to Figure 4.4, although the similarity score of “without the MSET” was similar, the score of “with the MSET” was higher on the gaze data than on the toy data. This difference might be because of the misdetection of attribute-of-focus due to the difference of the parameter of neutral-browsing model and durations of the focused browsing phase. In contrast to the actual gaze data, whose binomial parameter of neutral-browsing model was $p \in \{1/3, 2/3\}$ and the ratio of focused browsing phase was around 65% (as shown in Figure 3.9 (a)), we simulated with the toy data the browsing situation where the binomial parameter was $p \in \{1/4, 1/2, 3/4\}$ for all attribute types and the ratio of focused browsing phases was 33%. Therefore, the possibility of false positive from the MSET on the toy data expected to be higher than that of the actual gaze data.

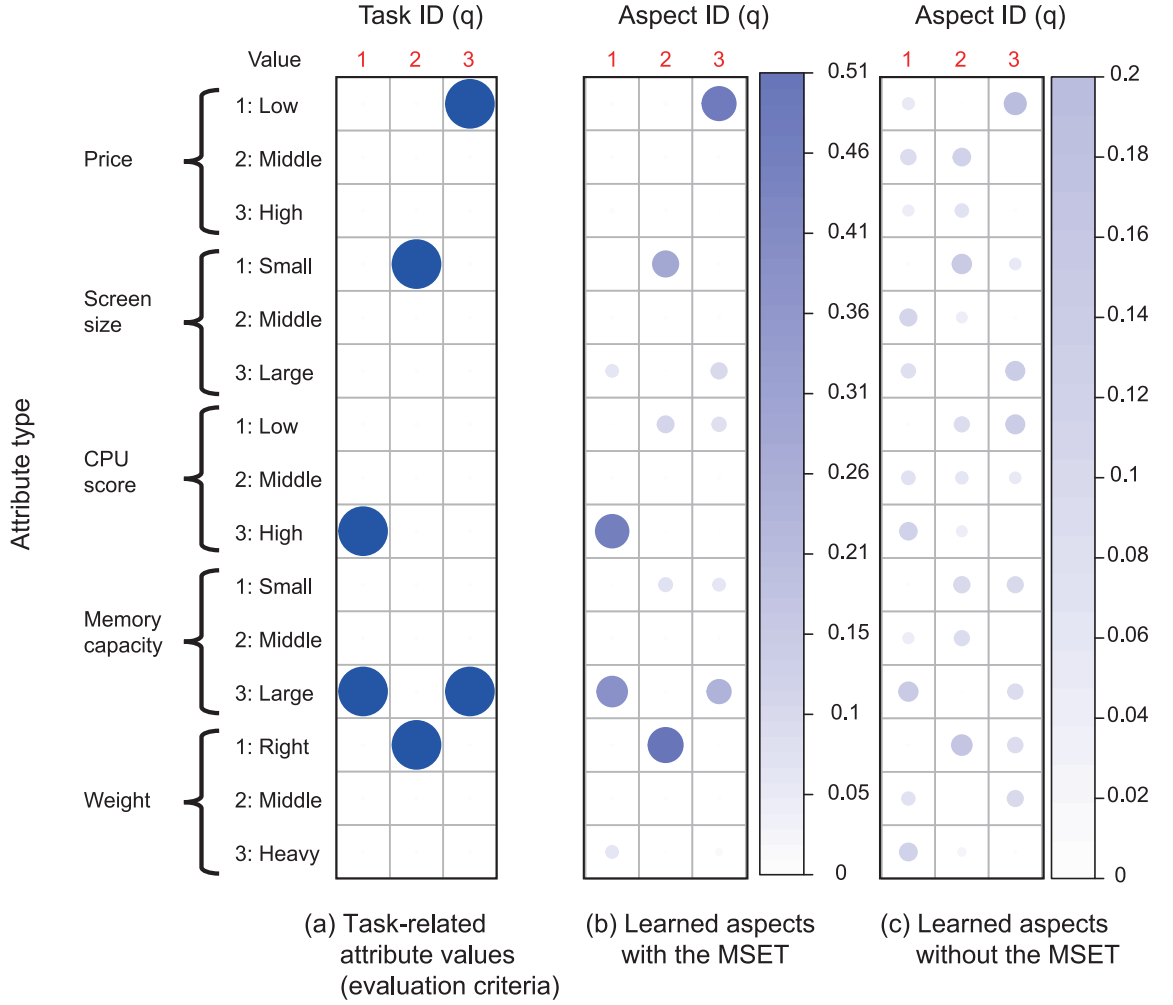


Figure 4.6: An example of task-related attribute values (given evaluation criteria) and learned aspects with and without the MSET on the gaze data. The number of aspects was set to be three, the same as that of tasks. The size and color of each dot depict the value of the multinomial parameter $p_{q,k}$, which represents the degree the aspect C_q is related to attribute values V_k .

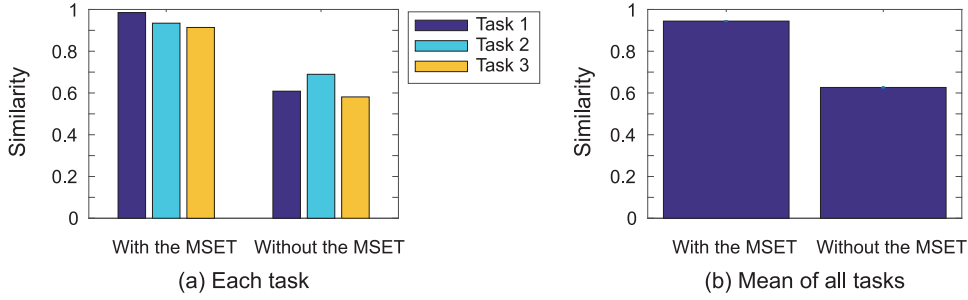


Figure 4.7: Similarities between task-related attribute values and aspects with and without the MSET. Each bar of (a) shows the similarity between learned aspect and each evaluation criterion given as the task on actual gaze data. (b) shows the mean of similarities of all tasks.

Table 4.2 shows the results of task estimation based on the maximum probability of the estimated selection interests, $\arg \max_q P(C_q|s)$ similar to Section 4.6.1. In 11 out of 111 sessions, the MSET did not detect users' comparison behavior as a bias of gaze behavior, and the accuracy of task estimation was 78.4%. The reason why the MSET could not detect participants' comparison behavior can be contributed to the duration lengths of content browsing in several sessions were quite short as shown in Figure 3.8. That is, participants decided on one item after looking just a few times and therefore did not compare items well.

Note that, in this experiment, we designed the tasks so that an attribute value was related to two evaluation criteria (i.e., high CPU score was an attribute value related to both task 1 and task 3) to confirm that participants' engaging tasks are distinguished by not only a single attribute value but the combination of several attributes. Although this design decreases the separability of the estimation of task 1 and task 3, the accuracy of the tasks were still higher than 80% once attribute-of-focus was detected by the MSET.

These results show the effectiveness of the proposed two-step approach, especially using the MSET, to learn evaluation criteria as the aspects of items and to estimate users' engaging task from participants' gaze behavior. In particular, the results of Figures 4.4 and 4.6 indicate that considering users' "focus" on attributes (i.e., attribute-of-focus) is crucial to analyze users' comparison behavior.

Table 4.2: Results of evaluation criterion estimation on gaze data with the MSET. Number in each cell shows the number of sessions that were estimated as engaged task q ($q = 1, 2, 3$). “No comparison” means the MSET could not detect a bias of gaze behavior in the session.

		Estimated task			
		1	2	3	No comparison
Actual task	1	28	0	5	4
	2	0	30	3	4
	3	5	0	29	3

Table 4.3: Results of evaluation criterion estimation on gaze data without the MSET. Number in each cell shows the number of sessions that were estimated as engaged task q ($q = 1, 2, 3$).

		Estimated task		
		1	2	3
Actual task	1	36	0	1
	2	2	29	6
	3	19	0	18

4.7 Discussion

In this section, we discuss the limitations of the proposed framework from the viewpoint of user's internal model.

Dynamics of users' interests Users' selection interests are affected by many factors such as individual preferences and temporary interests elicited by external information. In some research fields including education, these two different types of interests are often called *individual interests* and *situational interests* [Wentzel and Wigfield, 2009], respectively, and we borrow these terms in the following discussion. While users' individual interests do not depend on only intrinsic personality but the users' experience and can change over time, they are comparatively stable. On the other hand, situational interests change dynamically depending on stimuli (e.g., novel information, goals) provided to the users.

However, our model assumes that users' selection interests, $\{\theta\}$, are constant during content browsing, and in our experiments, we therefore gave the participants a task for each session. Using this experimental setting with tasks, we assume that the influence of situational interests was much more dominant than individual interests and that the selection interests were constant during each session. Although this experimental setting is preferable to verify the feasibility of the proposed model and algorithm for learning evaluation criteria, which is the main focus of this paper, users' selection interests change dynamically in actual situations and it should be addressed in order to put the method to practical use.

One approach is to exploit the results of attribute-of-focus detection by the MSET (Chapter 3). By dividing a sequence of attribute-of-focus with its change points of attribute values, the users' selection interests can be considered to be constant in each of the segments. Then, the proposed learning and estimation methods can be applied to these segments, i.e., $\{\theta\}$ is assumed to be piece-wise constant.

The degree of users' attention As explained in Section 4.4, we take account of users' attention resources in the proposed gaze-behavior model with the model parameter n_t , which is the number of attribute values of focus in a multinomial distribution $h(f_t; n_t, p_q)$. However, in the context of modeling attention resources, the degree of attention is considered; that is, a user allocates continuous-valued attention resources toward each object within a limited capac-

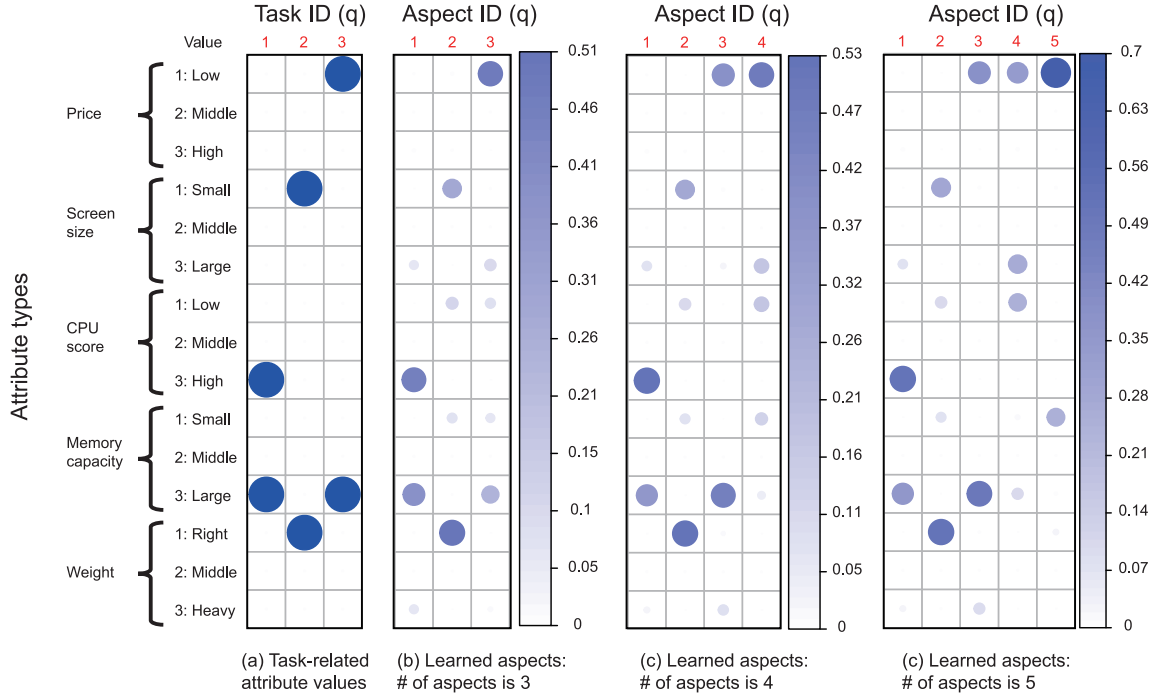


Figure 4.8: Learned aspects by changing the number of aspects

ity [Goldstein et al., 2011] in contrast to our model that described attention as a binary entity. Specifically, we use threshold-based analyses for both the identification of biased periods and the detection of attribute-of-focus in those periods.

While the binary representation of attention is simple, it poses another question how to determine appropriate thresholds and there may have some information lost through the binarization. For example, a significance level is used as the threshold for p -values but the MSET analysis completely ignores the periods whose p -value exceed the significance level as we discussed in Section 3.5. A simple remedy is to map p -values to the degree of attention to have $f_{t,k} \in [0, 1]$ instead of thresholding, which yields $f_{t,k} \in \{0, 1\}$.

The number of aspects The proposed method finds evaluation criteria based on unsupervised learning likewise topic models, and therefore the number of aspects needs to be given. In our experiments, the given number of aspects was the same as that of tasks to evaluate the learned aspects in terms of the similarity to the given tasks. In practice, the number of evaluation criteria can vary depending on users and situations (e.g., content and tasks) as mentioned in Section 2.4.1.

Figure 4.8 shows an example of learned aspects with different number of aspects. In this result, aspects C_1 to C_3 are determined by the sum of the similarities

of three pairs that became the largest among all possible pairs similar to the experiment. Aspect C_4 in the results of the number of aspects is 4 and 5 shows that “cheaper, large screen size, low CPU score” is one aspect of participants’ decision making. This aspect can be interpreted as “To watch movie with large screen.” Because the task in the experiment was not explicitly mentioned as a set of several attribute values, several participants may focus on the screen size to watch the movies. However, “cheaper” attribute is strongly related to aspect C_4 and C_5 when the number of aspects is 5. Therefore, one possible reason is that participants sometimes focus on the “cheaper” alternatives and the other attributes are happen to detected by the MSET.

To determine the appropriate number of aspects, one approach is to apply standard hyper-parameter estimation of unsupervised learning, e.g., non-parametric Bayes. However, the “interpretation” of learned aspects is desirable for many applications, such as speech dialog systems using evaluation criteria for probing questions [Misu et al., 2011]. While this issue is not covered in this study, a procedure for determining the number of aspects needs to be investigated in terms of the interpretation of learned aspects.

Chapter 5

Gaze Behavior Model with the Spatial Effect for Item Recommendation

5.1 Introduction

In this chapter, we introduce the effect of the spatial layout of the content into our proposed gaze behavior model. As mentioned in Chapter 1, users' gaze behavior is affected by not only users' internal states but also the visual attention to the content. For example, users tend to look at a salient region [Itti et al., 1998] or around the center of the content (center bias) [Borji, 2012]. Also, the degree of the effect of these external factor depends on the situation, content, users and can be changed. Therefore, how to discriminate these two effects is one question to estimate users' internal states.

The idea of our approach is that both internal factors (i.e., selection interests) and external factors (i.e., visual attention) are dealt with in a unified model. To unify these effects, the observation of the model is a region of the content, in contrast to the model in Chapter 4, whose observations are attribute-of-focus. Therefore, this model can generate not only on which attribute values users focus but where users look at, that is, gaze behavior.

In this study, we specifically focus on the absolute position in a content i.e., center bias, and do not consider other effects (i.e., salient region or temporal relations of gaze targets).

The contribution of this chapter is as follows: We introduce the effect of the

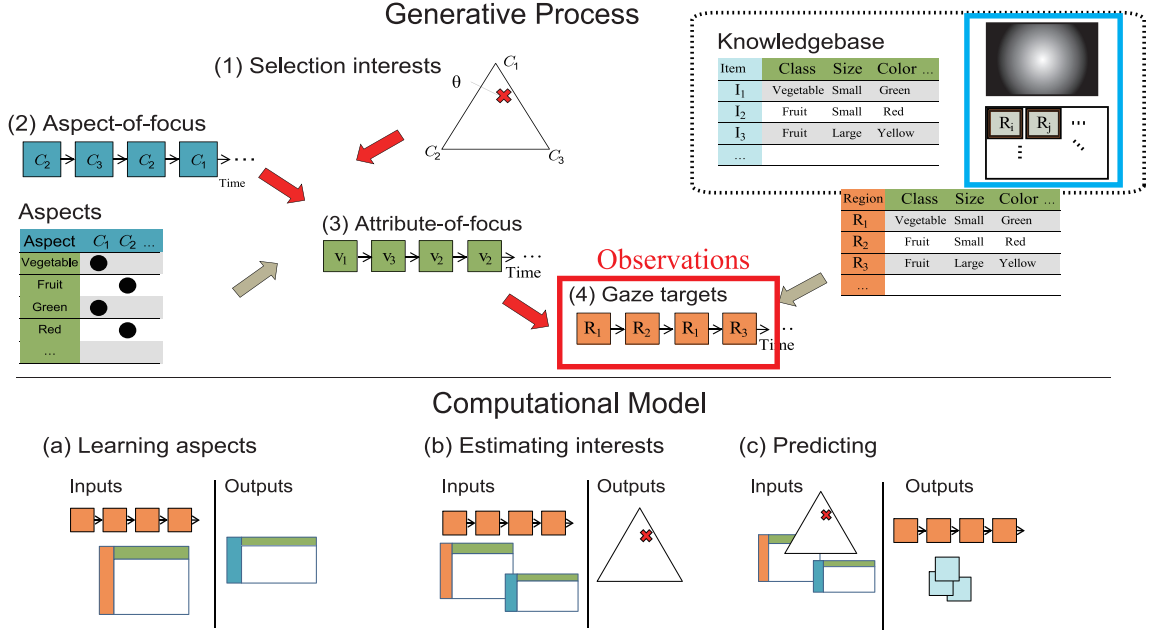


Figure 5.1: Generative process of gaze behavior from users' selection interests

spatial layout to users' gaze behavior model. The proposed model can estimate users' selection interest and predict where users tend to look at or which alternatives are preferred by users based on estimated selection interests by taking spatial layout into account.

5.2 Generative Process of Gaze Behavior

As mentioned above, the observations of this model $\mathcal{O}$ is the region that occupied alternatives $\{R_{n_t}\}_t$. As denoted in Section 2.5, these regions $R_n \in \mathcal{R}$ is corresponded to each of item $I_n \in \mathcal{I}$.

The generative process is as follows: At first, as introduced in Chapter 2, users' aspect-of-focus $c_t \in \mathcal{C}$ is determined by users' selection interests. Then, attribute-of-focus $v_t \in \mathcal{V}$ is determined according to the probability distribution $P(v_t = V_k | c_t)$. In this chapter, we simply assume that users focus on one attribute value in each time, and therefore, this probability distribution $P(v_t = V_k | c_t)$ is a categorical distribution. Note that, each aspect $C_q \in \mathcal{C}$ is modeled by the degree of association with attribute value V_k (i.e., $P(V_k | C_q)$). Finally, region $r_t \in \mathcal{R}$ is looked at depending on displayed item $I(r_t)$ with the attribute-of-focus v_t (as described below, we model this as $P(r_t | v_t)$).

Here, we assume that conditional probabilities $P(r_t = R_n | v_t = V_k)$ and $P(v_t =$

$V_k|c_t = C_q$) are constant. Then, the joint probability of a region-of-gaze, attribute-of-focus and aspect-of-focus for given session s is derived as follows:

$$P(r_t = R_n, v_t = V_k, c_t = C_q|s) = P(R_n|V_k)P(V_k|C_q)P(C_q|s). \quad (5.1)$$

Hence, the probability that the user looks at the region R_n is given by

$$\begin{aligned} P(r_t = R_n|s) &= \sum_{k=1}^K \sum_{q=1}^Q P(R_n, V_k, C_q|s) \\ &= \sum_{k=1}^K \sum_{q=1}^Q P(R_n|V_k)P(V_k|C_q)P(C_q|s) \end{aligned} \quad (5.2)$$

by marginalization over V and C in Equation (5.1). As a result, the probability of the region-of-gaze sequence in session s , $\{R_{n_t}^s\}_t = (R_{n_1}^s, \dots, R_{n_{T_s}}^s)$, is derived as

$$P(\{R_{n_t}^s\}_t) = \prod_{t=1}^{T_s} P(R_{n_t}^s|s) = \prod_{t=1}^{T_s} \left\{ \sum_{k=1}^K \sum_{q=1}^Q P(R_{n_t}^s|V_k)P(V_k|C_q)P(C_q|s) \right\}. \quad (5.3)$$

5.3 The Effect of Layout in the Gaze Behavior Model

Since a gaze distribution, which is a distribution of alternatives in a sequence of users' gaze targets, on a display are affected by not only the user's selection interest but also the position of items and its spatial design, some existing research introduced a prior knowledge that users tend to look at the center of the display. Indeed, the frequency distribution of region-of-gaze through all sessions in our experiment (described in detail in Section 5.6) is given as Figure 5.2, and we can see that the center of the display tends to be looked at than surrounding areas. In this distribution, not the actual center but upper of the center tends to be looked at. This may because the participants' head is located an upper side of the content.

In this chapter, we try to deal with the gaze behavior by taking the effect of the content layout into account. As introduced in Chapter 2, the equation of our model (Equation (5.1)) is based on the pLSA, which is one of the probabilistic generative models. Figure 5.3 shows a graphical model of the pLSA and proposed model which the effect of the layout. Although the pLSA cannot deal with the spatial and temporal structures among words (i.e., observations), in terms of the eye movement analysis, these structures of gaze targets are important (e.g., item

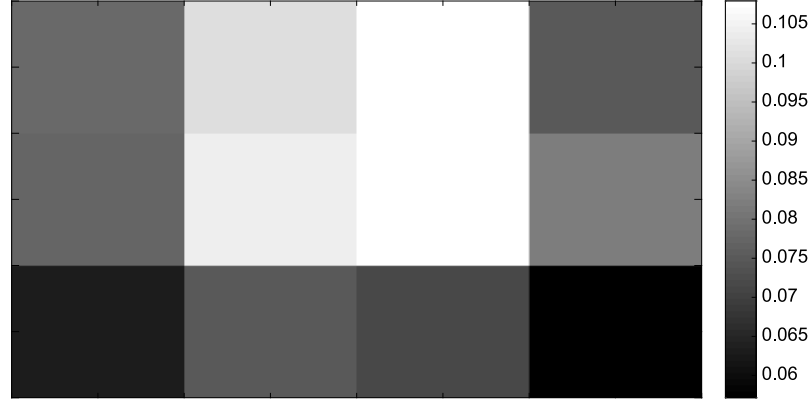
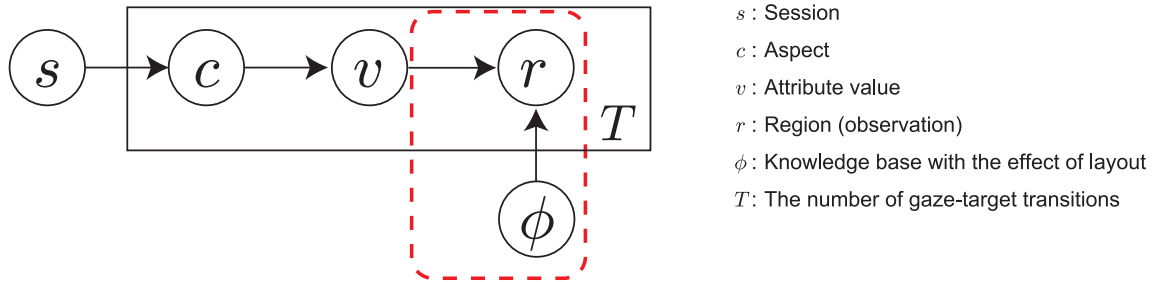


Figure 5.2: Normalized frequency distribution of region-of-gaze on the display obtained from the experiment in Section 5.6

layout in the content and the temporal order of items that are looked at). Therefore, in this chapter, we extend the pLSA (Figure 5.3 (a)) as shown in Figure 5.3 (b).



(a) probabiliscit Latent Semantic Analysis (pLSA)



(b) Proposed model

Figure 5.3: Graphical model of the pLSA and the proposed model

In the pLSA, word w is generated from aspect z as an observation. In our model, in contrast, attribute-of-focus v , which is not observed from outside, is first generated from aspect c , and users' region-of-gaze, which is actually observed, is then generated from the attribute-of-focus v . Because this region-of-gaze r is affected by a content layout, we introduce probability distribution ϕ when users determine region-of-gaze r from attribute-of-focus v (red-dotted frame in Fig-

ure 5.3 (b)). This ϕ is a mixture of the effect of the layout and a knowledge base of the content domain, which contains what attributes does each item in the content have. We will discuss in detail in Section 5.5 how to consider the effect of layout and how to model ϕ . By adding one layer between observations and aspects, the conditional probability $P(R_n|V_k)$ can be calculated in advance, and aspects can be learned and user's selection interests can be estimated independently the model of ϕ .

As such, considering $P(R_n|V_k)$ instead of $P(I_n|V_k)$ directly is important in the formulation in the previous section. And, this enable us to use the structure of content layout. Although $P(R_n|V_k)$ also depends on ϕ , in what follows we denote as $P(R_n|V_k)$ for simplicity.

5.4 Estimation of Model Parameters

In this section, we list the parameters of estimation by taking the effect of layout into consideration. At first, we suppose that the probability of regions conditioned by each attribute value, $\{(R_n|V_k)\}_{n,k}$, is known in every content.

Here, the probability of the region-of-gaze sequence in session s is given by Equation (5.3). Therefore, the probability of the set of region-of-gaze sequences $\mathcal{O}$ is given by

$$P(\mathcal{O}) = \prod_{s=1}^S \prod_{t=1}^{T_s} \sum_{k=1}^K \sum_{q=1}^Q P(R_{n_t}^s|V_k)P(V_k|C_q)P(C_q|s). \quad (5.4)$$

That is, optimization equation is described as follows:

$$\begin{aligned} & \underset{\{\theta(s)\}_s, \{p_q\}_q}{\text{maximize}} && \sum_{s=1}^S \sum_{t=1}^{T_s} \log \left(\sum_{k=1}^K \sum_{q=1}^Q P(R_{n_t}|V_k)P(V_k|C_q)P(C_q|s) \right) \\ & \text{subject to} && \sum_q \theta_q(s) = 1 \quad \forall s \\ & && \sum_k p_{q,k} = 1 \quad \forall q. \end{aligned} \quad (5.5)$$

In this chapter, gaze target region R_t is regarded as the observation of the model. Therefore, the prediction is "which region the user tends to look at." However, in terms of the prediction, "which item the user tends to take selection interest in" can also be predicted without considering the effect of layout

once aspects are learned. That is, both

$$P(R_n|s) = \sum_{k=1}^K \sum_{q=1}^Q P(R_n|V_k; \phi) P(V_k|C_q) P(C_q|s), \quad (5.6)$$

and

$$P(I_n|s) = \sum_{k=1}^K \sum_{q=1}^Q P(I_n|V_k) P(V_k|C_q) P(C_q|s) \quad (5.7)$$

can be predicted by this generative model.

5.5 Model of Effect of Layout

In this section, we propose two ways of modeling ϕ in Figure 5.3, which is a probability distribution taking into account the effect of layout together with knowledge base, as described in the previous section. While region-of-gaze may have some temporal correlation in general, i.e., users tend to look at a region close to the previous gaze target, we particularly focus on the effect of the absolute position in a content as mentioned in Chapter 2.

5.5.1 Mixture Model

First, we propose a mixture (additive) model. In this model, we assume that a user has two latent states when the user looks at a region in a content. One state represents that the user focuses on items based on the user's selection interests, and the other state represents that the user looks at regions depending on only the spatial arrangement (i.e., layout). For example, the user tends to look at an item located close to the center of a content (so called a center bias). To combine these two states into the model, we here assume that a user browses a content by switching the two states.

In particular, we model $P(R_n|V_k)$ as follows. Suppose that we have knowledge base $KB = \{(I_n, V_k) \mid I_n \text{ has } V_k\}$, as the relation between item I_n and attribute value V_k . Then, we assume that the joint probability distribution of these two parameters is simply given by $P(I_m, V_k) = 1/|KB|$, where $|\cdot|$ denotes the cardinality of a set. Under this assumption, conditional probability $P(I_n|V_k)$, which is a probability of item I_n been interested in when attribute value V_k is given for attribute-of-focus, can be considered as follows, with the number of items, N , in

the displayed content:

$$P(I_n|V_k) = \begin{cases} \frac{P(V_k, I_n)}{P(V_k)} = \frac{P(V_k, I_n)}{\sum_n P(V_k, I_n)} & (V_k \text{ exists in the content}) \\ \frac{1}{N} & (\text{otherwise}). \end{cases} \quad (5.8)$$

We assume that when a user chooses an item of gaze, the user first recalls an attribute value related to the user's selection interest, and then chooses an item that has the recalled attribute value. Here, in case that no item in the content has the recalled attribute value, we assume that the user chooses one item with equal probability.

Then, we assume that the effect of the layout is given as the probability distribution $P(R_n; \gamma)$ with parameter γ , which is independent of displayed items. We finally model the probability of region R_n being looked at under given attribute value V_k as follows:

$$P(R_n|V_k) := (1 - \beta_t)P(I(R_n)|V_k) + \beta_t P(R_n; \gamma), \quad (5.9)$$

where β_t is a parameter that determines the mixture weights for time t .

5.5.2 Prior-Based Model

Another way to model the effect of the layout is to consider $P(R_n; \gamma)$ as a prior. An existing study in the field of visual saliency also proposed a method to calculate the saliency of the image by dealing with the effect of layout, center bias, as a prior [Borji, 2012]. In the previous section, we assume that a user has two states in content browsing, i.e., interest-driven focus on items and the effect of the layout are exclusive at a time. In contrast, we can introduce another model in which one state is affected or modulated by the other. In other words, we here consider that the two factors, i.e., selection interests and layout, simultaneously affect user's gaze distribution.

In particular, we model $P(R_n|V_k)$ as follows. Suppose we have two parameters $P(V_k, I_n)$ and $P(R_n; \gamma)$ as described in the previous section. Note that, we have relation

$$P(R_n|V_k) = \frac{P(V_k|R_n)P(R_n; \gamma)}{\sum_n P(V_k|R_n)P(R_n; \gamma)} \quad (5.10)$$

and that it is reasonable to assume $P(V_k|R_n) = P(V_k|I(R_n))$. Therefore, we model the probability distribution of region R_n being looked at given an attribute value

V_k as follows:

$$P(R_n|V_k) = \begin{cases} \frac{P(V_k|I(R_n))P(R_n;\gamma)}{\sum_n P(V_k|I(R_n))P(R_n;\gamma)} \propto \frac{P(V_k, I(R_n))}{\sum_q P(V_k, I(R_n))} P(R_n; \gamma) & (V_k \text{ exists in the content}) \\ P(R_n; \gamma) & (\text{otherwise}). \end{cases} \quad (5.11)$$

5.5.3 Difference of Two Models

Both these models consider the effect of layout (i.e., center bias), and can represent the difference of the probability of items looked at even though the items have same attribute values.

The mixture model is similar to smooth distribution $P(R_n|V_k)$ by mixture of $P(R_n; \gamma)$, therefore, while the estimation result of the prior-based model is sensitive to prepared attributes, that of the mixture model can be generalized by smoothing. Moreover, since the prior-based model assumes two factors, interest-driven focus on items and the effect of layout, simultaneously affect user's gaze distribution, this model cannot consider that the region happens to attract the user's gaze while the user changes gaze target to others.

Although the mixture model assumes that two states of content browsing are exclusive, the parameters themselves are calculated by marginalizing these two states. As a result, while the mixture model and the prior-based model use different calculation of $P(R_n|V_k)$, i.e., additive and multiplicative, respectively, both models seem not so different in terms of approximating the distribution of $P(R_n|V_k)$ as will be discussed in the next section. In addition, if the same parameter γ is used, the probability distribution can be tuned by β_t for the mixture model. Hence, the use of the mixture model, Equation (5.9), has an advantage in terms that the model is able to adaptively change the weight of interest-driven and layout-driven gaze behavior by using the parameter β_t .

5.6 Evaluation

In order to evaluate the proposed model, we conducted experiments with recorded gaze data when participants were browsing displayed contents. To verify whether a variety of aspects can be represented by the proposed approach, we give participants one task in each session so that they are encouraged to browse a content from various viewpoints. We also assume that giving tasks enables par-

ticipants to keep their selection interests constant in one session. Since this model generates the region-of-interests, we evaluate this model in terms that how well this model simulates users' gaze behavior, in contrast to the evaluation in Chapter 4, which is from the point of aspect learning. That is, the evaluation is how the model can predict users' gaze-of-interests and items-of-interests through selection interests estimation. In this chapter, we also evaluate our model from the points of the difference of attributes and the number of aspects. To do that, semantic information of the content is not presenting in the content, and not only semantic attributes but also appearance attributes are prepared for the model parameters.

5.6.1 Experimental Setting

Participants Nine participants (5 males and 4 females, ranging from age 20 to 30) took part in the experiment. The order of tasks was randomized for each participant, and the order of displayed contents in one task was also randomized. Every participant had normal vision or corrected vision. Since the used eye tracker could not measure eye gaze direction correctly if the participant put glasses, participants were allowed to correct their vision only by contact lens.

Equipment A display¹ and an eye tracker² under the display connected to a personal computer were installed. Each of participants was asked to sit in front of the display and to browse contents on the display. Eye movements of each participant were captured by an eye tracker.

Displayed contents and tasks Food images were used for contents, where each of the contents had a tiled layout of the images (as shown in Figure 5.4). We got permissions to use these images for our experiments and prepared five image sets consisting of twelve images. Every image was originally used for web pages of recipe³, or book of recipe⁴. Two out of the five image sets were used for all tasks (i.e., three tasks), and the remaining three image sets were used for each task (see Table 5.1). In order to avoid the effect of specific layout for one item, the

¹DELL, 24inch, W520mm/H325mm, resolution:1920×1200pixel

²Tobii X120 Eye Tracker, freedom of head movement is 300×220×300mm, sampling rate is 60Hz, accuracy is 0.5 degrees.

³Mizkan Co., Ltd. (<http://www.mizkan.co.jp/index.html>)

⁴Sally, photo by Takeshi Kubota (2013). I scientificized "delicious" and made it a recipe (in Japanese). Sunmark.

Table 5.1: Image set used in each task

Image set ID	1	2	3	4	5
Task 1	○	○	○		
Task 2	○	○		○	
Task 3	○	○			○

order of laid out images in one content was randomized for each session. Hence, although all participants took part in the same tasks with the same image sets, each displayed content including their spatial layout was different.



Figure 5.4: Example of displayed content. Every image is diverted from a web page, Mizkan Co., Ltd. (<http://www.mizkan.co.jp/index.html>), and a book (Sally, photo by Takeshi Kubota (2013). I scientificized “delicious” and made it a recipe (in Japanese). Sunmark.) by getting permissions for using in our experiments.

Each participant was engaged in the following three tasks:

- Choose three foods in the order you think are healthy.
- Choose three foods in the order you want to eat when you are hungry.
- Choose three foods in the order you want to cook by yourself.

In each of the above tasks, an ordered list of items was asked to be answered in the expectation that each participant compared and explored more interested items.

5.6. EVALUATION

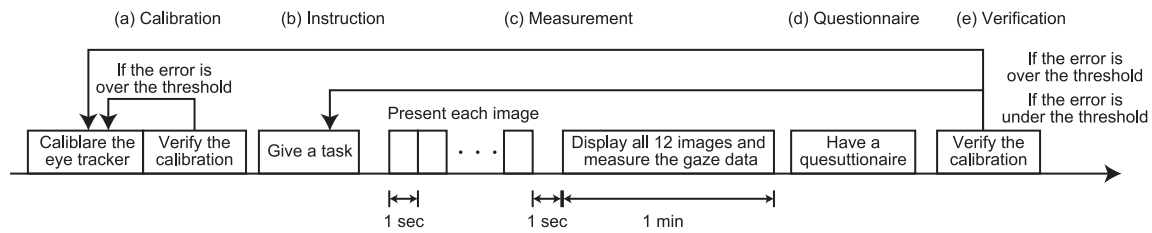


Figure 5.5: The procedure of the experiment

5.6.2 Procedure

Each of the participants was first asked to sit in front of the display. Here, the position of the seat was determined individually so that the participant's eye gaze direction could be obtained correctly by the eye tracker (about one meter away from the display). In this experiment, a chin rest was not used since the situation we assume was free-viewing of digital signages.

The procedure for the experiments was as follows (Figure 5.5): We first (a) calibrated the eye tracker, and then for each session, (b) gave a task, (c) presented a content, (d) had a questionnaire, (e) verified the calibration. Each participant was engaged in three tasks as described in the previous section, and in each task, the participant took three sessions with changing the displayed content. As a result, each participant took nine sessions. Therefore, the number of sessions in this experiment is $9 \text{ participants} \times 3 \text{ tasks} \times 3 \text{ contents} = 81 \text{ sessions}$.

For the calibration (Figure 5.5 (a)), five points were used to calibrate the parameters of the eye tracker and the other four points were used to measure the calibration error similar to Chapter 3. Then, the participant was explained a task (Figure 5.5 (b)) and a content was displayed (Figure 5.5 (c)).

Similar to the procedure in Chapter 3, to explicitly distinguish users' glimpse phase from comparison behavior in decision making [Russo and Leclerc, 1994, Shi et al., 2013], all items were displayed at intervals of one second in turn before the measurement of gaze data for analysis. Thanks to this phase, participants preliminarily receive all items in a content, and the bottom-up visual attention for items was expected to be decreased.

After this phase, all twelve images in the content were re-displayed all together, and the point-of-gaze sequence was captured starting from this time for one minute. After finishing one session, three items chosen in that session were noted by a questionnaire (Figure 5.5 (d)). This is because we also expect that the effect of the displayed content in a current session to the next session can be

decreased by having a questionnaire before the next session and not keeping in mind that items.

Then, the calibration error (Figure 5.5 (e)) was measured similar to (Figure 5.5 (a)). If this error was below a threshold, we proceeded to the next session; otherwise, we took a calibration phase (Figure 5.5 (a)) before proceeding to the next session.

5.6.3 Evaluation Scheme

In the evaluation, by assuming that a participant looks at and chooses items depending on their selection interests, our model was evaluated by the prediction accuracy of region-of-gaze that the participant tends to look at and item-of-interest that the participant selected by Equation (5.6) and Equation (5.7), respectively. Here, the prediction accuracy of items was evaluated by comparing to the items chosen by the participant in a questionnaire.

Note that, aspects are related to viewpoints from which participant looks at items in the content, and these aspects are to be learned for each content domain, e.g., food catalog, map, etc. Moreover, since a variety of viewpoints can appear when participants browse a content, it is preferred to prepare a wide range of attribute values in the learning process of aspects. Taking this into account, as for attributes in this experiment, food category (rice, meat, vegetable, fish and noodle), calorie and the number of steps to cook, were prepared as the semantic attributes. In particular, information of recipes placed together with the food images on the web pages or the book (see Section 5.6.1) was used. To reduce the number of attribute values, original calories were quantized into four clusters. On the other hand, color information in the entire region of each of the images was prepared as appearance attributes. Specifically, hue and intensity of images were calculated and vector quantized into sixteen clusters.

In order to evaluate the difference of these attributes, we conducted the analysis with the following three combinations:

SA: only the semantic attributes

AA: only the appearance attributes

SA+AA: both the semantic and appearance attributes

To verify how the number of aspects influences the results, two types of the number of aspects were prepared, and the prediction accuracies were compared.

As described in Chapter 2, we assume that participants' selection interests were fixed in one session. Therefore, participants' selection interests in one session were estimated from a part of the session. Here, a leave-one-out (LOO), cross-validation scheme was adopted. The calculation procedures of prediction accuracy of region-of-gaze and item-of-interest were the following.

At first, aspects were learned from the other 80 sessions. Also, the effect of layout, γ , were calculated by using gaze distributions in these 80 sessions. Then, the participant's selection interest was estimated from first $x[\%]$ of the test session using the learned aspects. Here, several values of x were adopted to see how the length of sessions used for the estimation affects the prediction accuracy.

Next, probability distributions for the regions and items in the content were estimated using estimated selection interest and learned aspects by Equation (5.6) and Equation (5.7). Here, these distributions serve as the "prediction" of the region-of-gaze of remaining $100 - x[\%]$ of the session and that of item-of-interest, respectively. That is, both region prediction (compared to actual region-of-gaze in remaining part of the session), and item item prediction (compared to three items chosen by the user) were evaluated. As for the measurement of the evaluation, the precision, which is often used in the field of information retrieval, was utilized.

The grand truth for regions and items were obtained by the actual region-of-gaze obtained from $100 - x[\%]$ of the session and by a questionnaire, respectively. Since three items chosen by the user were used for the grand truth of the item prediction, three regions that were most frequently looked at by the participant were used for the grand truth of the region prediction. In the calculation of precision, the order of the three regions and items were not considered.

Since we assume that the effect of the layout is constant as described in Section 5.5.1, the weight parameter β_t in the mixture model is fixed and the subscript t is omitted.

Baseline Method

Nearest neighbor (NN) recommendation was used as a comparison method, which is one of the basic methods in collaborative filtering and can predict item-of-interest. In order to accommodate with the proposed method, "what attributes participant focuses on" was used instead of "what items participant looks at" regarding using features. Specifically, the frequency distribution of region-of-gaze, $n(s, R_n)$, was calculated from point-of-gaze sequence in each session, and the fre-

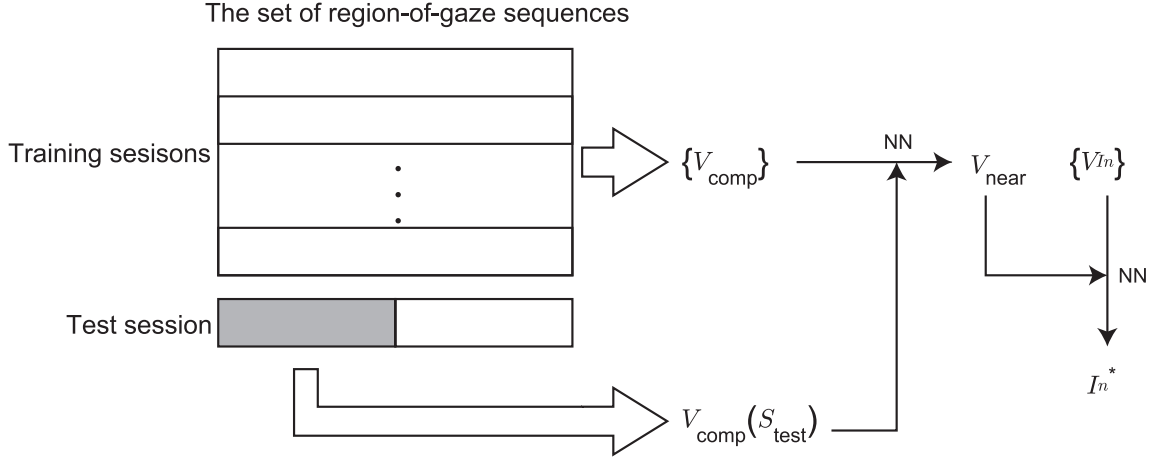


Figure 5.6: Flow chart of the comparison method

quency distribution of item-of-gaze, $n(s, I(R_n)) = n(s, R_n)$ was derived. Here, we re-denoted an attribute-value vector of each item as $V_{I_n} \in \{0, 1\}^K$ by using all possible attribute values. Each element of this vector denotes by 1 and 0 whether item I_n has the attribute value (1) or not (0). In other words, if an item has attribute value $v_k \in \mathcal{V}$, k -th element of the vector becomes 1. For each session, s , an attribute-value vector of interest, V_{comp} , was calculated using V_{I_n} , and the frequency distribution of item-of-gaze, $n(s, I_n)$, was calculated as follows:

$$V_{\text{comp}}(s) = \sum_{n=1}^N n(s, I_n) V_{I_n}. \quad (5.12)$$

Flow chart of this comparison method is shown in Figure 5.6. For each test of LOO, a set of attribute-value vectors of interest, $\{V_{\text{comp}}(s) \mid s \in \text{learning sessions}\}$, was prepared. Then, an estimated attribute-value vector $V_{\text{near}} \in \{V_{\text{comp}}\}$ was obtained, which had the nearest cosine distance from the attribute-value vector of interest $V_{\text{comp}}(s_{\text{test}})$ calculated from the first part of the test session. Finally, as a prediction for item-of-interest, one item was chosen whose attribute-value vector V_{I_n} was the nearest from the estimated attribute-value vector V_{near} . In addition, the region that contained the predicted item was obtained as a prediction of region-of-gaze.

Although this comparison method has a similarity to the proposed method in terms that it also considers attribute values that items have, this is inherently a memory-based method while the proposed method is model-based. Hence, comparing these two methods can verify the effectiveness of the modeling of gaze

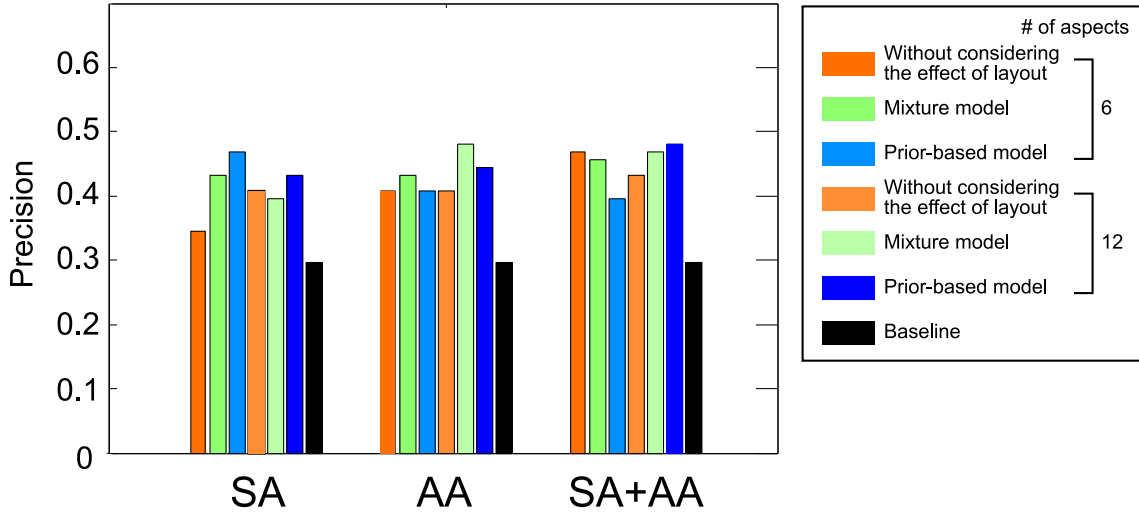


Figure 5.7: Prediction accuracy of region-of-gaze. 50[%] of each session was used to estimate selection interest.

behavior.

5.6.4 Results

Prediction of region-of-gaze

We show in Figure 5.7 the results of predicting region-of-gaze for two numbers of aspects ($Q = 6, 12$), respectively. Here, we set $\beta = 0.5$ as a mixture ratio in the mixture model described in Section 5.5.1.

Prediction accuracies were calculated for three types of combination of attributes, SA, AA and SA+AA as described in Section 5.6.3. In this result, 50[%] of each test session was used to estimate the selection interests. The vertical axis shows the obtained prediction accuracy by comparing predicted regions and actual region-of-gaze in the remaining 50[%] of each test session. In all situations, the proposed model can predict region-of-gaze with higher accuracy. We can also confirm that considering the effect of layout improves prediction accuracy.

Figure 5.8 shows the result of prediction accuracy of region-of-gaze with respect to different parameter β and the ratio of each test session used for the interest estimation (the number of aspects: 6, used attributes: SA). In this result, prediction accuracy fell when the ratio was 90[%]. This is because our evaluation scheme compared predicted regions and actual region-of-gaze in remaining $100 - x$ [%] of each test session. In short, the longer session the model uses for estimation, the shorter session the model can use for evaluation, and more difficult

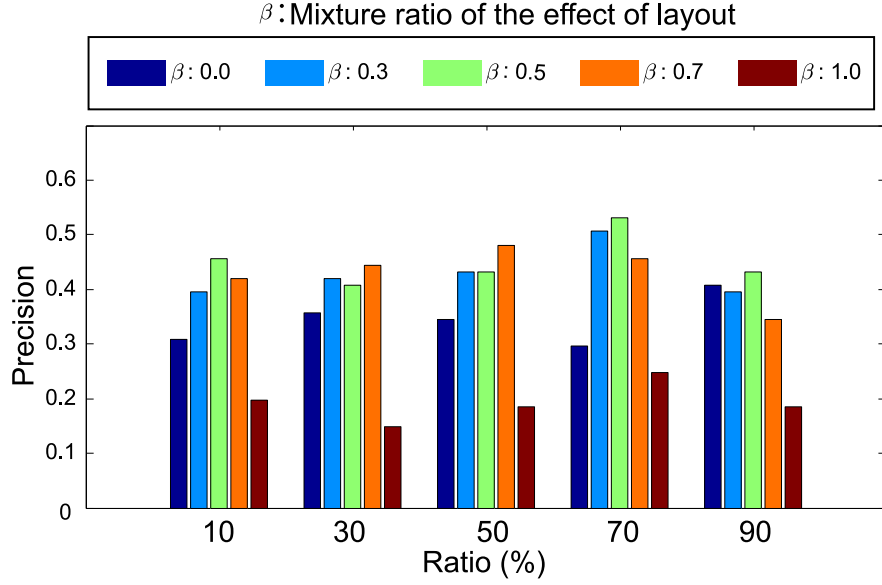


Figure 5.8: Prediction accuracy of region-of-gaze with respect to different mixture ratio β (the number of aspects: 6, used attributes: SA) and different ratio in each of test sessions used for estimating the participant's selection interests.

to predict region-of-gaze in these remaining sessions.

Prediction of Item-of-Interest

Figure 5.9 shows the results of predicting item-of-interest for two different numbers of aspects ($Q = 6, 12$), using $\beta = 0.5$. These results show that the proposed model can also predict item-of-interest of each session with higher accuracy compared to baseline method. We can also see that the prediction accuracy of the prior-based model often becomes very lower depending on using attributes.

Figure 5.10 shows the result of prediction accuracy of item-of-interest with respect to different parameter β and the ratio of each test session used for the interest estimation (the number of aspects: 6, and attribute: SA). This result shows that longer session the model uses for estimation, better prediction of item-of-interest the model can obtain.

Learned aspects

The aspects learned from all sessions for two different number of aspects ($Q = 3, 12$) are shown in Figure 5.11. In this figure, each column shows the relation between an aspect and attribute values, $P(V_k|C_q)$. Therefore an attribute value

5.6. EVALUATION

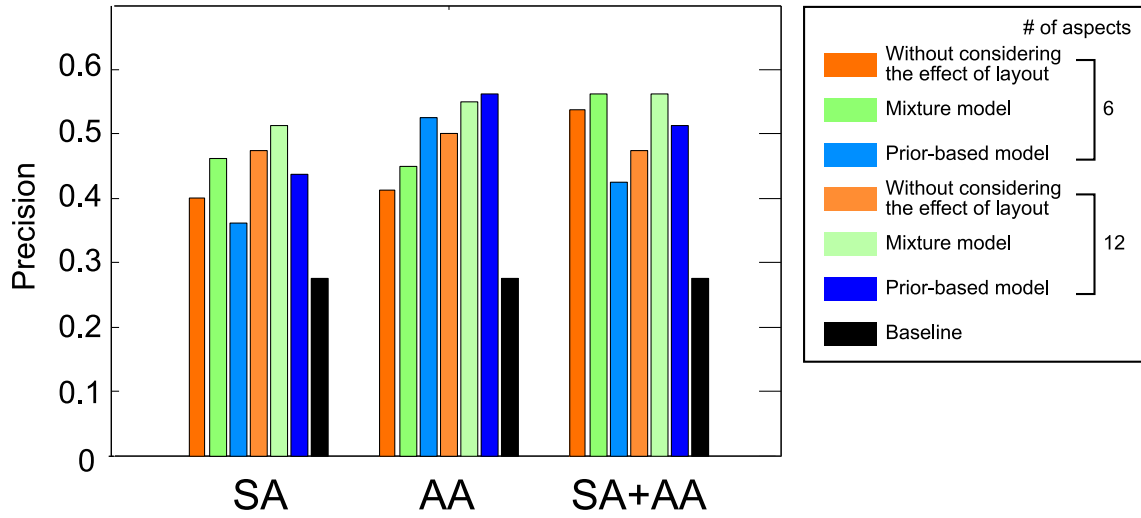


Figure 5.9: Prediction accuracy of item-of-interest. 50[%] of each session was used to estimate selection interest.

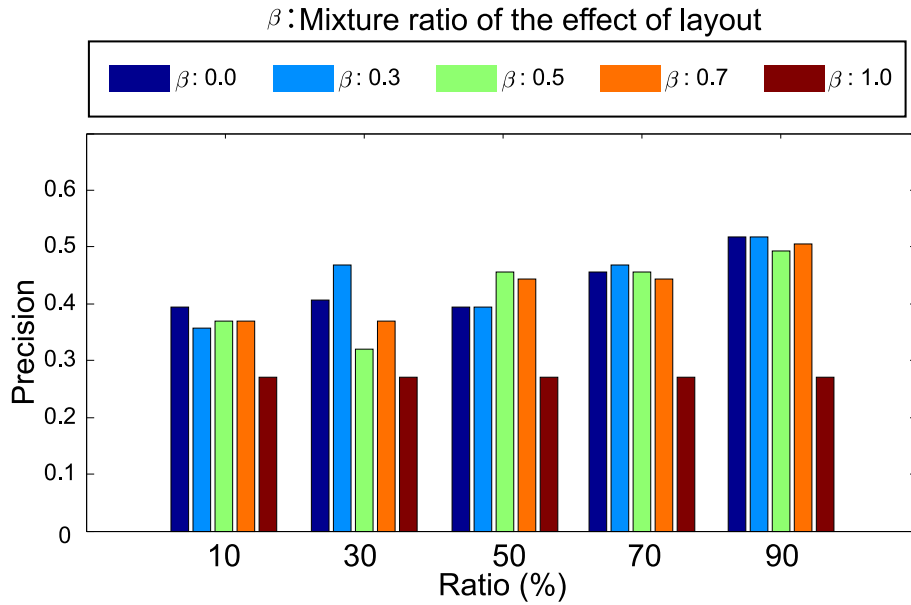


Figure 5.10: Prediction accuracy of item-of-interest with respect to different mixture ratio β (the number of aspects: 6, used attributes: SA) and different ratio in each of test sessions used for estimating the participant's selection interests.

with high score has more strong relation with that aspect. Here, only semantic attributes were used because each aspect was easily interpreted. In this analysis, the mixture ratio of the effect of the layout was $\beta = 0.5$.

In addition, estimated selection interests $\theta(s)$ in each session when browsing image set 1 are shown in Figure 5.12, where these selection interests are estimated

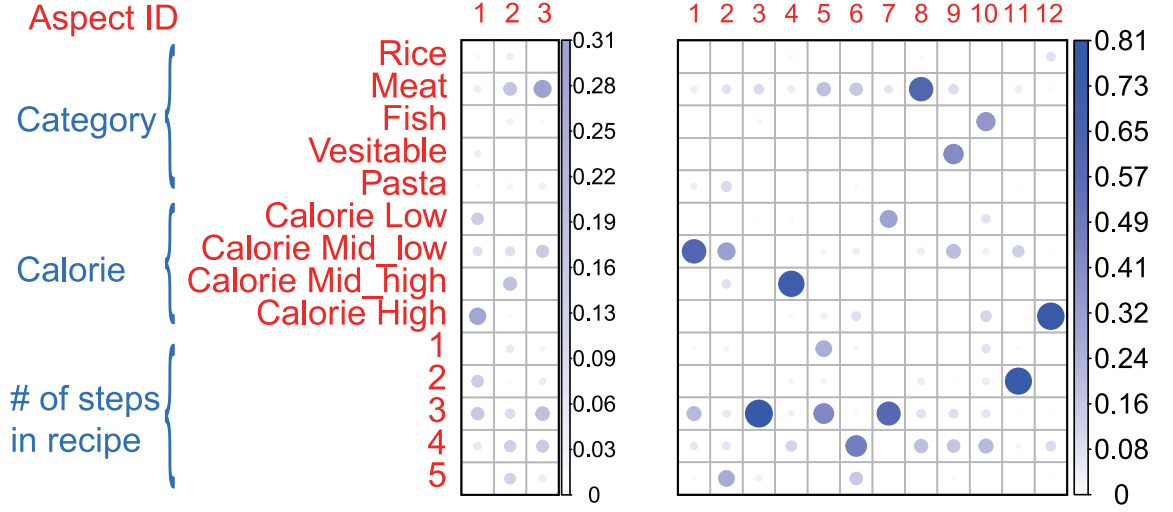


Figure 5.11: Examples of learned aspects ($\beta = 0.5$, used attributes: SA). Left: the number of aspects: 3, right: the number of aspects: 12. Attribute values with high score have strong relation with that aspect.

simultaneously with 12 aspects in Figure 5.11.

5.7 Discussion

Effectiveness of the proposed model From the results of Figures 5.7 and 5.9, we observe that the proposed method successfully predicts user's region-of-gaze and item-of-interest even though the training data set is not large enough. This result indicates the effectiveness of the proposed gaze behavior modeling since the model is considered to have contributed on the generalization of selection interests from a small size of observation.

Effectiveness of the considering the effect of layout Regarding the effect of layout, we showed that by including this effect into the model (Sections 5.5.1, and 5.5.2), the prediction accuracy can be improved (Figures 5.7 and 5.9). We can see that an appropriate β exists in $(0, 1)$, and the most appropriate value varies depending on the timing of estimation in a session (Figures 5.8 and 5.10).

This result also suggests that the value of β_t should be changed over time while constant value was used during a session of the experiments. While all prediction accuracy of item-of-interest is improved in the mixture model, some prediction accuracy of region-of-gaze becomes worse. This result may be caused

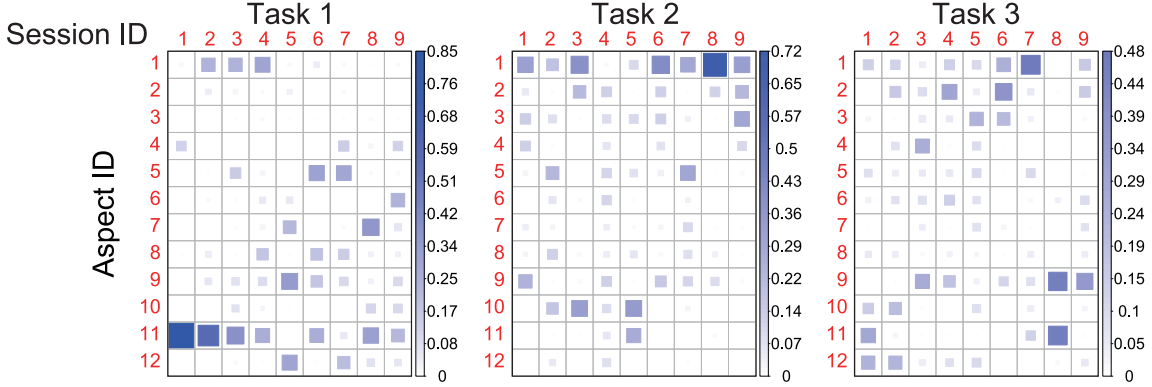


Figure 5.12: Examples of estimated selection interests $\theta(s)$ in each session ($\beta = 0.5$, used attributes: SA, displayed image set: 1). Sessions with high score have strong relation with that aspect shown in Figure 5.11 (right).

by constant β_t , i.e., even though a user compares items of interests in the latter part of the session, the proposed model predicts regions with the effect of layout same as the first part of the session.

Difference of the use of attribute types Figure 5.7 shows that the use of different attribute types does not strongly affect prediction accuracy in the mixture model. Therefore, we can obtain a good prediction of region-of-gaze only using the appearance attributes that can be extracted automatically by image processing in contrast that the semantic attributes require manual annotation. Note that, if we want to interpret learned aspects, it is useful to use the semantic attributes. On the other hand, Figure 5.9 shows that using both the semantic and the appearance attributes yields the highest accuracy, and we can say that it is better to use a various types of attributes.

Difference of two models By comparing the prior-based model to the model without considering the effect of layout, we observe that the prediction accuracy of both region-of-gaze and item-of-interest sometimes takes lower values. As discussed in Section 5.5.3, the prior-based model assumes that when a user looks at an item, the user's attribute-of-focus is one of the attribute value that items have. Therefore, we consider that the prior-based model is more affected by prepared attributes compared to the mixture model.

Interpretation of learned aspects Regarding the interpretation of learned aspects, the use of less number of aspects ($Q = 3$) associates each aspect with more attribute values (Figure 5.11, left). On the other hand, the use of larger number of aspects ($Q = 12$) makes attribute values associated with each aspect sparse, and it is easy to be interpreted (Figure 5.11, right). For example, the 6-th aspect can be interpreted to “elaboration,” which requires many steps for cooking, and the 9-th aspect can be related to “healthy,” since it has large weights on vegetable and low-calorie (Figure 5.11, right).

Difference of the number of aspects In general, more aspects we set, more interpretable results we obtain because few number of attribute values are related to each of aspects as shown in Figure 5.11. However, in terms of prediction accuracy, increasing the number of aspects does not necessarily improve the accuracy. For instance, when we predict region-of-gaze with the semantic attributes (Figure 5.7), using 12 aspects has lower accuracy than using 6 aspects due to over fitting of the model. In summary, we have to choose an appropriate number of aspects depending on the purpose of the model.

Representing variety of user’s selection interests From the estimated selection interests of each session when browsing image set 1 shown in Figure 5.12, we observe that different tasks cause different aspects. While task 1 and task 2 have some specific peaks in the estimated distributions, task 3 (choose three foods in the order you want to cook by yourself) involves various types of aspect distributions (selection interests). This result indicates that task 3 had more degree of freedom for the users to choose items. Even in such a case, the proposed model seems to successfully deal with user’s different selection interests in individual sessions.

Chapter 6

GazeAHP: Tracing Selection Interests and Supporting Users to Recognize the Structure of Problems

6.1 Introduction

The analytic hierarchy process (AHP) [Saaty, 2008] is one of the basic structured techniques to support users' MCDM as mentioned in Section 1.2.3. Users understand the problem by structuring the hierarchy. Once the hierarchy is constructed, the utilities of each alternative can be calculated, and the users can organize or recognize their selection interests. However, if users do not have enough knowledge about the domain, structuring a hierarchy becomes difficult.

The use of interactive recommender systems is one approach to this issue. If the structure of decision problems of concern is given as a hierarchy, the system can assist users' decision making by suggesting evaluation criteria. For example, Misu et al. [Misu et al., 2011] applied the hierarchical structure of a problem to a user model and proposed a spoken dialog system to assist users' decision making by interactively providing knowledge about the content domain through verbal interaction. Since this system estimates user preferences by pairs of system's action and users' reaction, a large number of interactions are required for the estimation, which may degrade smoothness of decision making.

In this chapter, we propose a framework named *GazeAHP*, which is an interactive recommendation based on users' gaze behavior. Given a hierarchy of decision problem with evaluation criteria, the system is able to estimate user se-

lection interests from users' gaze behavior instead of users' pairwise comparison as shown in Chapter 2.

To interactively assist users' decision making, following questions need to be addressed:

- How the system determines the action to take.
- How the system determines the action timing.

In this study, we adopt a multiscale detection approach proposed in Chapter 3 to decide an action timing. This approach detects an attribute value on which users specifically focus, which we refer to as *attribute-of-focus*, by short-term analysis. Also, by preparing the relationship between aspects and each of attribute values, an action is determined by estimated selection interests using the method introduced in Chapter 4.

The aim of interactive assistance is to assist users to understand the structure of the problem and construct a selection interests. Therefore, we try to estimate how easy users' selection interests can be changed, i.e., ambiguity of users' selection interests, while selection interests estimation approach introduced in Chapter 2 is a point estimation by maximum likelihood estimation. By estimating both users' selection interests and its ambiguity, the system can change strategies of actions. For example, if the user's selection interests can easily change, i.e., the ambiguity is large, the system suggests an evaluation criterion to assist users to construct selection interests while the system recommends items based on current selection interests when the ambiguity is small. By assuming Dirichlet distribution $P(\theta)$ for the probability distribution of selection interests θ , users' selection interests and its ambiguity can be estimated by a Dirichlet parameter α .

We show the effectiveness of the GazeAHP in terms how the system's interaction helps users "understand the problem of decision making and choose items" by proactively providing the knowledge of association between user preferences and attributes of items and how the users' selection interests changes as a reaction to the system's action.

6.2 The framework of GazeAHP

The procedure of the AHP is described in Section 1.2.3.

1. Structure the decision problem as a hierarchy.

2. Evaluate elements in the hierarchy.
3. Calculate utilities of alternatives.
4. Decide one alternative.

In the GazeAHP, an interactive system has a given hierarchical structure of the problem as shown in Figure 2.3, and the system estimates user preferences on the hierarchy. In the GazeAHP, the system estimates user preferences as the importance weight on each of evaluation criteria from gaze behavior instead of step 2. Once the system estimates user preferences, the system probes user preferences by simultaneously presenting an evaluation criterion according to the estimated preference and attribute values of items which users look at. This probing is crucial for decision support because users can structure the problem into a hierarchy as step 1 by the information of the association between evaluation criteria and attribute values provided by the probing. Besides, according to the utilities of each element in the structure provided by the system (step 3), the users can easily evaluate alternatives against evaluation criteria and decide an alternative (step 4).

6.3 Interaction design

Interactions for decision support consists of two steps: *probing* and *sorting*. In this chapter, we hypothesize that the proactive assistance that provides users the information of association between evaluation criteria and attribute-of-focus helps users understand the problem of decision making. Therefore, not only the attribute-of-focus V_t^f detected by the MSET but also the estimated users' aspect-of-focus C^f are simultaneously shown to the users during the interaction. Here, attribute-of-focus is first detected as the attribute value on which the user specifically focus while multiple attribute values were detected as attribute-of-focus in Chapter 3. Then, users' aspect-of-focus is estimated instead of step 2 of the AHP as the evaluation criteria based on gaze behavior model introduced in Chapter 4 as $C^f = \arg \max_q P(C_q | V_t^f)$.

As we will describe in Section 6.4, the temporal change of selection interests with its ambiguity is also estimated for post-analysis. However, the strategy of interactive assistance uses only estimated aspect-of-focus because the main target here is to confirm the effect of above two basic actions.

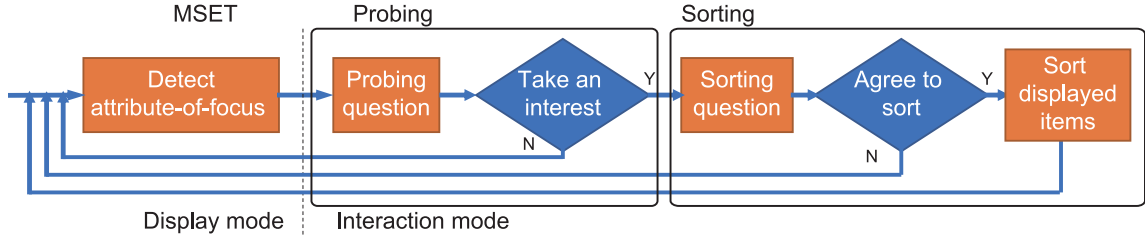


Figure 6.1: Interaction flow. Orange squares show the system’s action and blue diamond shapes shows the participants’ reaction.

Probing step The system probes users by asking whether estimated aspect-of-focus is collect or not using detected attribute-of-focus such as “You are browsing items with *green color (attribute value)*. Are you interested in *healthy food (evaluation criteria)*?” With the confirmation of the system’s estimation to the users, the system can update estimated user states and avoid taking incorrect actions. Besides, the system’s explicit mentioning of estimated aspect-of-focus and detected attribute-of-focus facilitates the users to structure the problem (step 1) as mentioned in the previous section.

Sorting step This step corresponds to steps 3 and 4 of the AHP introduced in Section 6.2. In this step, the system rearranges items based on user preferences (step 3) so as to assist users’ comparison behavior. Before the system rearranges the content, the system explicitly asks users whether the users want to sort or not in order to avoid users’ confusion to the sudden change of the content and to share the evaluation criteria for the sort. This sharing enables the users to evaluate alternatives on the preference and helps users decide alternatives (step 4). For example, “If you are interested in *healthy food (evaluation criteria)*, *vegetable and green color (attribute values)* may be the important criteria. Do you want to sort the displayed content based on *healthy food (evaluation criteria)*?”

The flow of the GazeAHP is shown in Figure 6.1. The system first probes user preferences, and then, if the system can confirm that the estimation is correct, the system takes sorting action.

6.4 Tracing Temporal Changes of Selection Interests

While selection interests estimation introduced in Chapter 2 adopts a maximum-likelihood estimation, i.e., point estimation, for simplicity, the posterior probability distribution of selection interests itself can be estimated (i.e., with Bayes

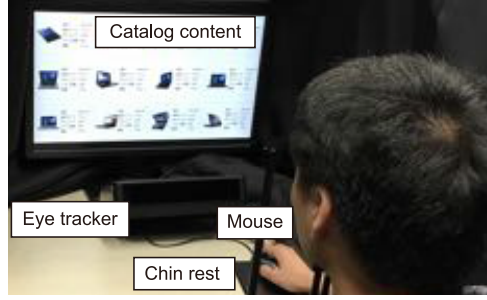


Figure 6.2: Experiment environment

estimation) since the proposed model is a probabilistic generative model. As we mentioned in Section 6.1, in this chapter, by assuming a Dirichlet distribution for a probability distribution of users' selection interests, users' selection interests and its ambiguity are simultaneously estimated.

At first, a sequence of attribute-of-focus are obtained from users' gaze behavior by the MSET introduced in Chapter 3. Then, fixed-length window are applied to the sequence of attribute-of-focus, and estimate posterior probability distribution $P(\theta|f_w)$ for each of windows as follows:

$$\begin{aligned}
 P(\theta|f_w) &\propto \sum_q P(f_w|C_q)P(C_q|\theta)P(\theta) \\
 &= \sum_q h(f_w; n_w, p_q) \theta_q \alpha_{w,q},
 \end{aligned} \tag{6.1}$$

where $n_w = \sum_{t \in w} \sum_k f_{t,k}$, and p_q is a multinomial parameter corresponding to the aspect C_q as described in Chapter 4. Here, we assume uniform distribution for the prior distribution of the $P(\theta)$ (i.e., $\alpha_{w,q} = 1$). Finally, the Dirichlet parameter α_w , which generates a probability distribution $P(\theta)$, are estimated for each of windows based on [Minka, 2000]. As a result, the temporal changes of user's selection interests together with its ambiguity are traced as a sequence of α_w .

6.5 Evaluation

We conducted an experiment to evaluate the proposed interaction framework with the cooperation of 37 participants, who also participated the experiments in Chapters 3 and 4.

6.5.1 Experimental Design

In the experiments, participants were asked to select one item (laptop computer) from 12 displayed items on a screen (see Figure 6.2), and participants' eye movements were captured by an eye tracker¹. Each item had five attribute types such as CPU score and memory capacity.

The purpose of this experiment is to confirm the effect of providing the knowledge of the evaluation criteria as being related to attribute values by objective and subjective evaluation.

Therefore, two types of interaction strategies were prepared: *attribute-value based interaction* (AB) and *evaluation-criteria based interaction* (CB). Participants were divided into two groups to assign each of those two strategies. Three evaluation criteria C_q ($q = 1, 2, 3$) related to attribute values $\mathcal{V}_q^c$ were prepared for CB strategy same as the experiment in Chapter 3 (see Table 3.1). Here, by giving the same weights on each attribute values in Table 3.1 (i.e., $P(V_k|C_q) = 0.5, V_k \in \mathcal{V}_q^V$), users' aspect-of-focus were estimated by $\arg \max_q P(C_q|V_t^f)$. Although an attribute value "large memory capacity" equally relates to two aspects (i.e., C_1 and C_3) in this table, aspect C_1 , "To play games or watch movies at home," is estimated as an aspect-of-focus when "large memory capacity" is detected as an attribute-of-focus in this study.

Two modes, *display modes* and *interaction modes*, were prepared during interactive decision assists. In the display mode, the system only displayed items, and participants compared items based on individual preference. A timing of switching these modes is determined by the MSET. In particular, once the system detected participants' attribute-of-focus by the MSET, the system changed its modes from display mode to interaction mode. To explicitly switch the participants' attention between the two modes, the system showed dialog boxes, which completely hid the displayed content. Besides, the systems' questions (i.e., probing question and sorting question) were restricted to only yes-no question, and participants' answers were given by mouse clicks.

The interaction mode consisted of two steps as mentioned in Section 6.3. In contrast to the CB strategy, which simultaneously showed aspect-of-focus and attribute-of-focus to participants, the system showed only attribute-of-focus for the AB strategy.

¹Tobii X120 Eye Tracker: Freedom of head movement is $300 \times 220 \times 300$ mm, sampling rate is 60 Hz, and accuracy is 0.5 degrees.

Note that, several attribute values were not related to any evaluation criteria, which we refer to as *criteria-unrelated attributes*, since the number of prepared evaluation criteria was small. Therefore, if criteria-unrelated attributes were detected as attribute-of-focus, the system could not estimate users' aspect-of-focus and it was impossible to act to participants. To uniform the action timing among two strategies, the system took actions to participants only when the detected attribute-of-focus was included in the criteria-related attributes, $V_t^f \in \cup_q \mathcal{V}_q^c$.

6.5.2 Procedure

Each participant was first asked to sit facing the display and to position their face to be aligned with the chin rest (Figure 6.2). Then, the eye tracker was calibrated. After the calibration, participants were explained the content and procedures of this experiment. In this experiments, the task was "Please select one laptop computer, which you actually want to buy and use."

The procedure consisted of the following four steps.

- Step 1** Participants were explained three evaluation criteria as example criteria.
- Step 2** The accuracy of eye tracker's calibrated parameter was confirmed.
- Step 3** Contents were displayed, and the participants were asked to choose one item with system's assistance.
- Step 4** Participants were asked to answer a questionnaire with five-point Likert scales after they chose an item.

6.5.3 Analysis and Discussion

In the proposed framework, the system took actions to participants only when the criteria-related attributes were detected as attribute-of-focus. Therefore, if the attribute values were not detected during decision making, the interaction did not occur. In this experiment, the significance level to detect participants' attribute-of-focus was set to 5%. Criteria-related attributes were detected as attribute-of-focus from 30 out of 37 participants' decision making (16 for AB strategy and 14 for CB strategy), i.e., 30 participants decided an item with the system's interaction. We used an objective and subjective measure to evaluate our framework from these 30 participants' decision making.

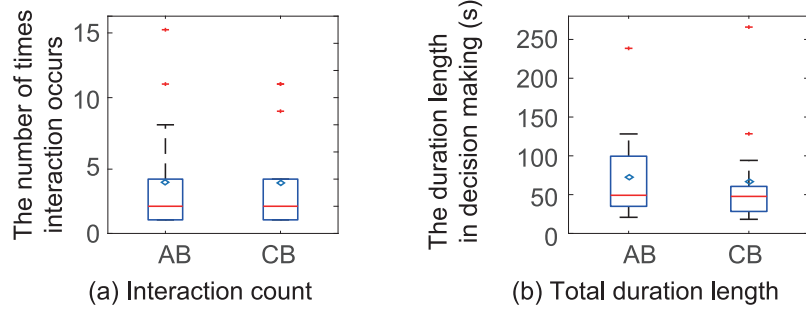


Figure 6.3: The number of times interaction occurs and the total duration of decision making

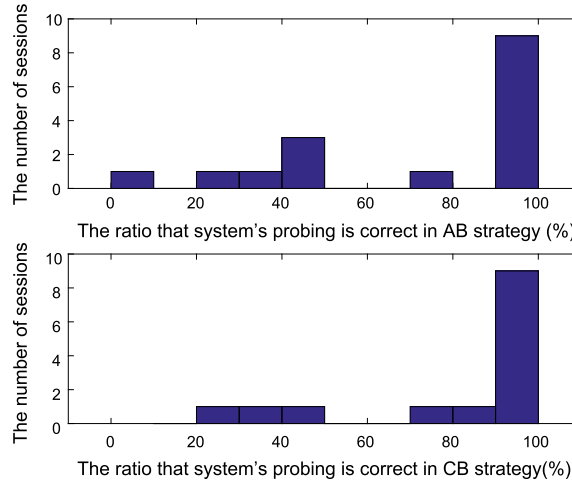


Figure 6.4: The ratio that the system's probing action is correct

Objective evaluation

Figure 6.3 shows the result of the number of times interaction occurred and the total duration length of decision making. The mean of the number of interaction times were 3.75 and 3.71, respectively, and the mean of the duration length of AB strategy and CB strategy were 72.5 and 66.5 sec., respectively.

Here, we show how the system's estimation is correct in Figure 6.4. The accuracy of the probing action is more than 70% and all the system's probing actions were true in more than half sessions, 9 of 16 sessions in AB strategy and 9 of 14 sessions in CB strategy. Since how easy participants can decide items depends on the ratio of the system's estimation, we picked the results only when all the system's probing are collect. Figure 6.5 shows the results of the participants' decision making that all the estimated selection interests were collect. The difference were not significant ($p = 0.34$ and 0.19 with Welch t-test, respectively) but these results indicate that participants with CB strategy could decide with shorter duration

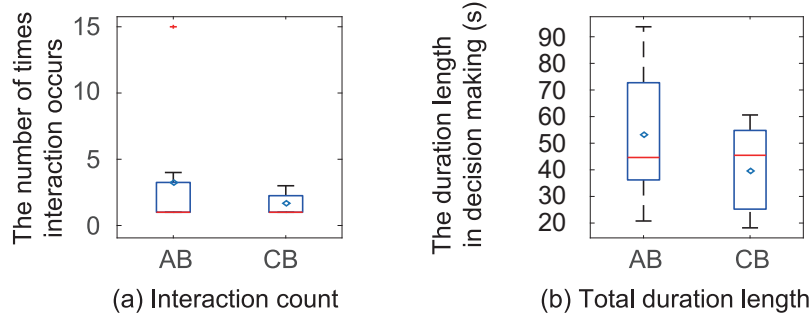


Figure 6.5: The number of times interaction occurs and the total duration of decision making only when all the estimated selection interests were collected during one choice behavior

length, and indicate that with the system’s probing by not only attribute-of-focus but also aspect-of-focus, the participants could smoothly decide items from alternatives.

Figure 6.6 shows examples of transition of the Dirichlet parameters in the session and the timing of interaction. In these results, multinomial parameter $\{p_r\}$ was given by evaluation criteria of given tasks similar to Chapter 4 with additive smoothing. These results show that the participant in CB strategy compares several alternatives from different aspects in turn just after sorting and then compare alternatives that satisfy the criteria, while the participant in AB strategy first glances at the content because of the re-arranging the content and compare alternatives that satisfy the criteria. The aim of sorting action is to assist participants comparison behavior based on an aspect-of-focus by re-arranging alternatives based on participants’ aspect-of-focus. However, this result implies that sorting action assists participants to once change aspect-of-focus and to browse content from different aspects by gathering alternatives having higher evaluation score, i.e., divergent action.

The histogram of the difference of mean of α_t^f are shown in Figure 6.7. These results show that almost all sorting action in AB strategy help participants decrease ambiguity of the selection interests. Also, while half of sorting actions in CB strategy assist participants’ selection interests construction, the other half actions triggers the changes of aspect-of-focus and gives an opportunity to explore the content.

To confirm this difference of the users’ browsing strategy, we observe the difference of histogram of region-of-gaze. Figure 6.8 shows a histogram of region-of-gaze before and after sorting action. Figure 6.8 (a) shows a center bias of this

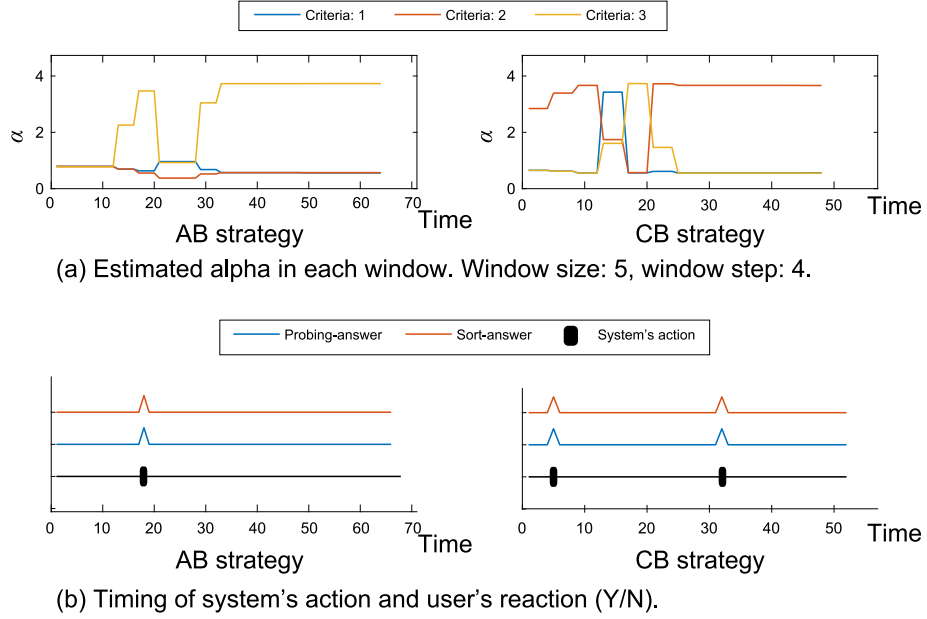


Figure 6.6: (a) An example of transition of estimated α_t and (b) the timing of system's action and participant's reaction, where + is yes, and - is no.

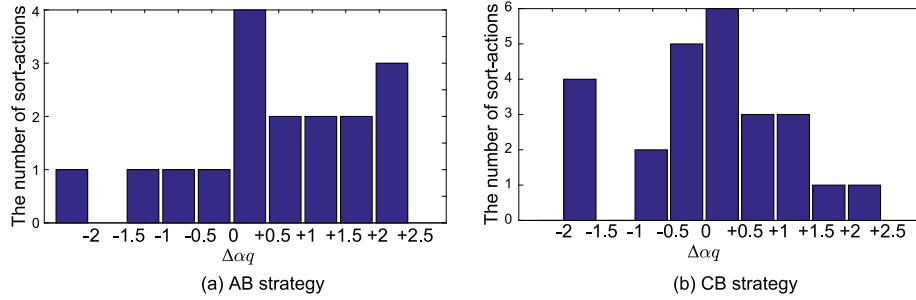


Figure 6.7: The histogram of transition of mean of α_q after system's action. (a) is a histogram of AB strategy and (b) is that of CB strategy.

content. The sorting action arranges items by total order and row-first. That is, the score of alternative located top-left is the highest and bottom-right is the lowest. These results show that participants in CB strategy well look at the alternatives having lower scores in terms of an aspect-of-focus (i.e., alternatives located in the bottom row), while participants in AB strategy look at the alternative having the highest score well (i.e., alternatives located in the top row).

Subjective evaluation

For participants' subjective evaluation, we asked participants the evaluations for the decision behavior including the decision process and chosen items, and eval-

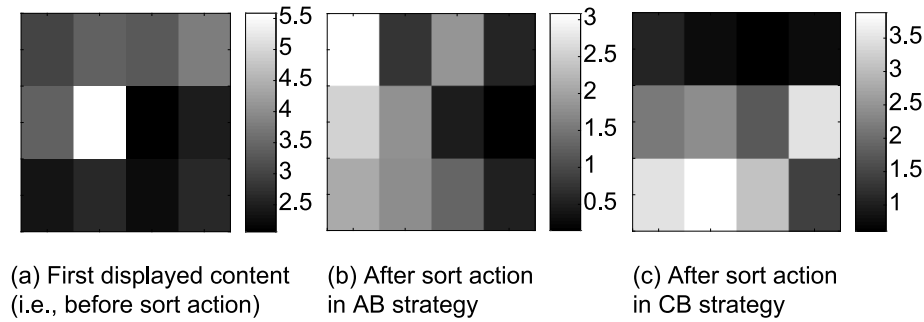


Figure 6.8: The histogram of region-of-gaze in the all sessions. (a) The histogram before sorting action (i.e., first displayed content) of all session and (b) the histogram after sorting action in AB strategy, and (c) the histogram after sorting in CB strategy.

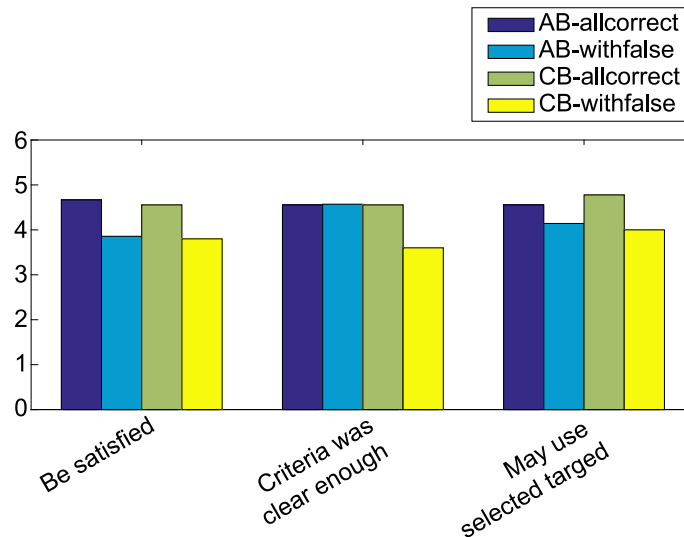


Figure 6.9: Subjective evaluation for the decision behavior

uations for the interactions between the system by the questionnaire.

Figure 6.9 shows the evaluation for the decision behavior conditioned by the ratio of probing action. Here, “allcorrect” means all probing actions are correct. Although we expected that participants with CB strategy tended to successfully construct their preference, and therefore those scores were higher than the scores of participants with AB strategy, those two results were not distinct. One possible reason is the prepared criteria were too simple. Therefore, participants could structure the decision problem as a hierarchy without the system’s suggestion. Moreover, the ambiguity of the criteria may be another reason. The evaluation for interactions with the system were shown in Figure 6.10. Almost all scores of evaluation of CB strategy were higher than that of AB strategy especially when

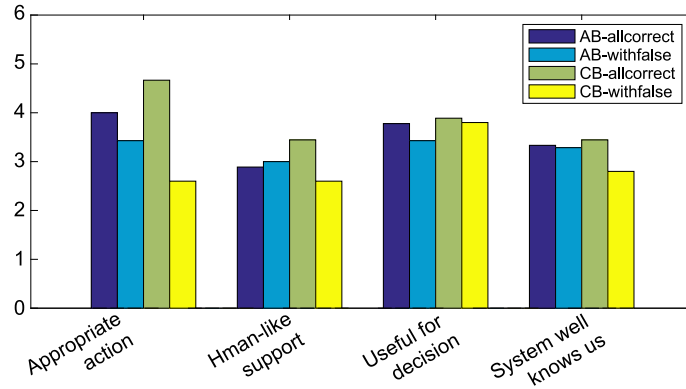


Figure 6.10: Subjective evaluation for interactions with the system

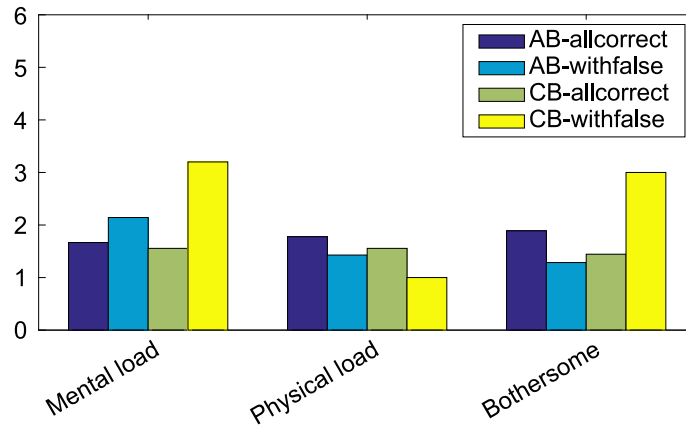


Figure 6.11: Subjective evaluation for load for the interaction

the system's estimation is correct. In particular, the results show that CB strategy was felt more "Appropriate action," the result of CB is 4.7 and AB is 4.0 while the significance is not large ($p = 0.10$ with Welch t-test). However, these results also show the fault of the system's action decreases the scores of evaluation for interaction. A possible reason of this result is the system's action in CB strategy mentions not only users' attribute-of-focus detected by the MSET but the estimated selection interests, and therefore participants feel bothered to the system when the estimation is failed. This is also confirmed in Figure 6.11. The score of "Mental load" and "Bothersome" are higher when the system's estimation is failed in CB strategy ($p = 0.18$ and 0.03 , with Welch t-test to "CB-withfalse" and "AB-withfalse", respectively).

Besides, we asked participants "why did you think this system is useful/useless for your decision making?" In qualitative feedback, nine of 16 participants with AB strategy commented "the sorting makes my decision easier."

6.5. EVALUATION

That is, in AB strategy, the sorting step had an impact on participants' decision making.

In contrast, in the case of the participants with CB strategy, six out of 14 participants also mentioned sorting step, but four participants commented "the system interacts me by the criteria on what I actually focus," and one participant commented "the system organizes information in my place." These results show that the impact of system's probing step became larger than AB strategy in the CB strategy and might increase the evaluations.

Chapter 7

Conclusion

7.1 Summary

In this thesis, we propose a novel modeling of users' selection interests and a framework for estimating users' selection interests from gaze behavior. The contributions of this thesis are: (1) We introduce aspects of items as users' evaluation criteria for decision making and introduce a novel representation of users' selection interests by the importance weight of aspects; and (2) we propose a generative model of users' gaze behavior by focusing on users' comparison behavior to learn aspects and to estimate users' selection interests from users' gaze behavior.

At first, we introduce a notion of aspects and a framework of the proposed gaze behavior model in Chapter 2. To model users' selection interests, we assume that aspects of items can be possible reasons for users' comparison behavior (i.e., why the user compares alternatives) and that these aspects can be characterized by the degree of association with each of attribute values. Under this assumption, the proposed framework learns aspects (i.e., the degree of association with attributes) from users' gaze behavior by maximum likelihood estimation. Gaze-behavior models proposed in Chapters 4 and 5 inherit the framework.

In Chapter 3, we propose a users' comparison detection method by introducing the statistical hypothesis test for multiple scales. By assuming users' neutral browsing behavior model, users' comparison behavior is detected by the deviation of gaze behavior distribution from neutral browsing model. Then, we propose a framework to learn users' possible evaluation criteria via a bottom-up method, which explicitly takes users' *focus* into account in Chapter 4 by concluding Chapter 3.

In Chapter 5, we take the effect of visual attention into account. In particular,

the absolute position in the content, which is called center bias, is to be the target of this study. We introduce two kinds of model of the effect of layout and compare these models regarding prediction accuracies of the generative models.

Finally, we introduce an interactive assistance based on Chapters 3 and 4. The interactive assistance is based on the AHP, and the system assists users several parts of the procedure of the AHP instead of the users, such as structuring problems and giving assessment scores to each of elements. The experimental results show that our approach proposed in this thesis can be a basis of the interactive assistance.

7.2 Future Work

In this section, we discuss the limitations of the proposed framework and extensions. Then, we discuss prospects to the interactive assistant system based on the proposed framework.

7.2.1 Limitations and Extensions

We discussed the limitations of the proposed framework from the viewpoints of aspect learning and observable gaze behavior.

Aspect learning

One limitation of aspects learning is caused by the number of observed data while this framework (i.e., parameter inference) assume a large amount of data.

Unprepared attributes We represent aspects by association with attribute values that presented in the content. That is, aspects that related to unprepared attribute values cannot be learned. Besides, to learn multiple attribute values as same aspects, these multiple attribute values need to be presented in the same content. To solve these issues, possible approaches are to prepare more variety of attributes and to deal with the combination of multiple attribute values as a new attribute value. However, these approaches cause the increase of the number of attribute values, and then becomes difficult to learn parameters.

Interpretation of aspects As discussed in Chapters 4 and 5, the interpretation of learned aspects is important issue for applying the proposed framework to inter-

active systems while the example of assistance proposed in Chapter 6 uses aspects prepared in a top-down manner. To address this issue, one extension is to apply interpretation approach to topic models [Mei et al., 2007, Chang et al., 2009]. Here, the ease of interpretation can be one measure for determining the number of aspects.

Features of gaze behavior

The proposed model in Chapter 2 simplified users' gaze behavior in many aspects, for example, by assuming users' focus on attribute values are determined at each time independently. We here discuss some limitations arise from the simplification of the observation of gaze-behavior model. While the limitations in the previous section can be overcome by the current model, several extensions need to overcome the limitations in this section.

Temporal patterns of browsing targets For the estimation of users' interests in decision making, temporal patterns of browsing targets convey useful information. For example, a user's "re-fixation" on a target indicates that the user specifically focuses on the target and compares it with the other alternatives [Schaffer et al., 2016]. Although these temporal patterns are indirectly considered in the MSET as the identification of a bias of users' gaze behavior in a short-term window, explicit modeling of gaze-transition patterns by a gaze-behavior model has a possibility of finding natural interpretation of users' gaze behavior.

Duration information Since we are particularly interested in users' comparison behavior, time t of the proposed gaze-behavior model was chosen as the change of the gaze targets (i.e., gaze-target transition). This is suitable to treat the number of times a target is looked at as the importance weight on the target. However, this approach cannot examine how carefully the user looks at that targets. Sugano et al. [Sugano et al., 2013] showed with Random Forests that the duration length is one of the most contributed features to estimate users' interests. Taking physical duration information into account may enable us to examine the degree of users' focus on a gaze target while browsing content.

7.2.2 Prospects to Interactive Decision Assistance

In this section, we describe prospects of interactive decision support based on our approach. Under the situation where users do not have enough knowledge or do not have a clear understanding of users' selection interests, proactive assistance can be useful. Here, let us consider a concierges' decision assistance, which is one of the popular proactive support.

Concierges interact with customers and estimate customers' internal states for appropriate assistance. The motivation of their interaction is having a clone of the customers' internal states to support customers in customers' shoes. They first try to know customers' internal states such as preference, intention, and knowledge about the content domain. To do that, by presenting information or suggesting some criteria, they incrementally probe the internal states by customers' reaction. After that, they recommend alternatives with its association with customers' internal states.

To realize concierge-like assistance, the system needs to have several functions, such as presenting information to users, asking a question to users and suggesting criteria to understand users' internal states deeply. In this study, we specifically focus on a function that the system helps users organize the problems and we show an example of interactive assistance based on gaze behavior named *GazeAHP* introduced in Chapter 6. The *GazeAHP* introduces two actions to assist users to understand the structure of the problem and to construct a selection interest: probing and sorting. To evaluate these two actions, the system's action set is very limited and the system's action strategy is deterministic, which uses only aspect-of-focus detected by the MSET proposed in Chapter 3.

Here, we introduce several extensions as follows to extend our framework to concierge-like assistance. While the *GazeAHP* assumes that users do not have a structure of problems but have knowledge about what attributes the item has and what the attribute is, users sometimes do not have enough knowledge about the meaning of attribute value. For example, users know an alternative has a high CPU score but do not know what CPU score is. Therefore, giving not only the relation between each element in the problem structure but the information about each element to users are also important assistance from the viewpoint of the information presentation.

Also, our proposed framework can estimate a variety of users' internal states such as the ambiguity of users' selection interests as shown in the analysis in

Chapter 6. Since useful assist can be changed depending on the ambiguity of users' selection interests, estimating ambiguity of users' selection interests enables the system to effectively change the support strategy. For example, if users have only fuzzy understanding of their preference, the system suggests several evaluation criteria while the system recommends alternatives that match users' selection interests when users' selection interests are almost constant. Precisely because systems can have much more information and obtain much more information from not only users' verbal behavior but also users' non-verbal behavior (e.g., gaze behavior and micro-facial expression), the systems are expected to effectively change the strategies and provide information to the users.

Over the past decade, a much amount of interactive systems were proposed not only in task-oriented interactive systems but non-task-oriented interactive systems [Misu et al., 2011, Yoshino and Kawahara, 2015, Hiraoka et al., 2016, Lowe et al., 2017, Liu and Lane, 2017]. In these communication systems' research, several kinds of approach to learning interactive strategies are applied such as reinforcement learning in Partially Observable Markov Decision Process [Misu et al., 2011, Yoshino and Kawahara, 2015, Hiraoka et al., 2016], and also by Neural Network techniques [Lowe et al., 2017, Liu and Lane, 2017]. By applying these techniques, we believe that the proposed framework can be a basis of interactive assistant systems.

List of Related Papers

Journal Paper

- Shimonishi, K., Ishikawa, E., Yonetani, R., Kawashima, H., and Matsuyama, T. (2014). Learning aspects of interest from gaze (in Japanese). *Transactions on Human Interface Society*, 16(2):103–114.

Conference Paper (International, Peer-reviewed)

- Shimonishi, K., Kawashima, H., Yonetani, R., Ishikawa, E., and Matsuyama, T. (2013). Learning aspects of interest from gaze. In *Proceedings of the 6th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction*, GazeIn '13, pages 41–44.
- Shimonishi, K., Kawashima, H., Schaffer, E., and Matsuyama, T. (2016). Tracing temporal changes of selection criteria from gaze information. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces*, IUI '16 Companion, pages 9–12.
- Shimonishi, K. (2016). Modeling user's decision process through gaze behavior. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16, pages 536–540.

Presentation

- Shimonishi, K., Kawashima, H., Yonetani, R., and Matsuyama, T. (2013). Estimation of aspects of interest from gaze behavior analysis (in Japanese). In *Technical Report of IEICE, PRMU2013-28*, pages 53–58.

- Shimonishi, K., Kawashima, H., and Matsuyama, T. (2015). Probabilistic gaze behavior modeling while browsing content (in Japanese). In *National Convention of IPSJ*, 5Z-03.
- Shimonishi, K., Kawashima, H., and Matsuyama, T. (2015). Probabilistic aspect-oriented gaze behavior modeling in visual content browsing. In *Meetings of Image Recognition and Understanding*, MIRU2015.
- Shimonishi, K., Kawashima, H., Schaffer, E., and Matsuyama, T. (2016). Tracing temporal changes of interests for interactive decision support systems (in Japanese). In *The 30th Annual Conference of the Japanese Society of Artificial Intelligence*, 2O4-OS-23a-2.
- Shimonishi, K. and Kawashima, H. (2017). Tracing temporal changes of users' selection interests during target selection from gaze behavior (in Japanese). In *Meetings of Image Recognition and Understanding*, MIRU2017.

Bibliography

- [Athukorala et al., 2016] Athukorala, K., Medlar, A., Oulasvirta, A., Jacucci, G., and Glowacka, D. (2016). Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces, IUI '16*, pages 359–369.
- [Ball, 1997] Ball, C. (1997). A comparison of single-step and multiple-step transition analyses of multiattribute decision strategies. *Organizational Behavior and Human Decision Processes*, 69(3):195 – 204.
- [Beach and Mitchell, 1978] Beach, L. R. and Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *The Academy of Management Review*, 3(3):439–449.
- [Bednarik et al., 2012] Bednarik, R., Vrzakova, H., and Hradis, M. (2012). What do you want to do next: A novel approach for intent prediction in gaze-based interaction. In *Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA '12*, pages 83–90.
- [Bhargava et al., 2015] Bhargava, P., Brdiczka, O., and Roberts, M. (2015). Unsupervised modeling of users’ interests from their Facebook profiles and activities. In *Proceedings of the 20th International Conference on Intelligent User Interfaces, IUI '15*, pages 191–201.
- [Bhushan and Rai, 2004] Bhushan, N. and Rai, K. (2004). *Strategic Decision Making: Applying the Analytic Hierarchy Process*. Springer London.
- [Blei et al., 2003] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Machine Learning Research*, 3:993–1022.
- [Borji, 2012] Borji, A. (2012). Boosting bottom-up and top-down visual features for saliency estimation. In *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 438–445.

-
- [Borji and Itti, 2013] Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207.
- [Brandherm et al., 2008] Brandherm, B., Prendinger, H., and Ishizuka, M. (2008). Dynamic Bayesian network based interest estimation for visual attentive presentation agents. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 1, AAMAS '08*, pages 191–198.
- [Burke, 2002] Burke, R. (2002). Interactive critiquing for catalog navigation in e-commerce. *Artificial Intelligence Review*, 18(3-4):245–267.
- [Chang et al., 2009] Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., and Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*, pages 288–296.
- [Chen et al., 2013] Chen, L., de Gemmis, M., Felfernig, A., Lops, P., Ricci, F., and Semeraro, G. (2013). Human decision making and recommender systems. *ACM Transactions on Interactive Intelligent Systems*, 3(3):17:1–17:7.
- [Chowdhury et al., 2011] Chowdhury, S., Gibb, F., and Landoni, M. (2011). Uncertainty in information seeking and retrieval: A study in an academic environment. *Information Processing & Management*, 47(2):157–175.
- [Cohen and Goldberg, 1970] Cohen, J. B. and Goldberg, M. E. (1970). The dissonance model in post-decision product evaluation. *Journal of Marketing Research*, 7(3):315–321.
- [Cole et al., 2013] Cole, M. J., Gwizdka, J., Liu, C., Belkin, N. J., and Zhang, X. (2013). Inferring user knowledge level from eye movement patterns. *Information Processing & Management*, 49(5):1075 – 1091.
- [Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximumlikelihood from incomplete data via the em algorithm. *the Royal Statistical Society, Series B*, 39:1–38.
- [Dijksterhuis and Nordgren, 2006] Dijksterhuis, A. and Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1(2):95–109.

- [Dodson et al., 2013] Dodson, T., Mattei, N., Guerin, J. T., and Goldsmith, J. (2013). An English-language argumentation interface for explanation generation with Markov decision processes in the domain of academic advising. *ACM Transactions on Interactive Intelligent Systems*, 3(3):18:1–18:30.
- [Ehrgott et al., 2010] Ehrgott, M., Figueira, J. R., and Greco, S. (2010). *Trends in Multiple Criteria Decision Analysis*, volume 142.
- [Eivazi and Bednarik, 2011] Eivazi, S. and Bednarik, R. (2011). Predicting problem-solving behavior and performance levels from visual attention data. In *Proceedings of 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at Intelligent User Interfaces*, pages 9–16.
- [Fishburn, 1970] Fishburn, P. C. (1970). *Utility theory for decision making*. Publications in operations research. J. Wiley.
- [Goldstein et al., 2011] Goldstein, E. B., Vanhorn, D., Francis, G., and Neath, I. (2011). *Cognitive psychology : connecting mind, research, and everyday experience*. Belmont, CA Wadsworth/Cengage Learning, 3rd edition. Previous ed.: 2008.
- [Hansson, 1994] Hansson, S. O. (1994). *Decision Theory: A Brief Introduction*. Royal Institute of Technology.
- [Häubl and Trifts, 2000] Häubl, G. and Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*, 19(1):4–21.
- [Hiraoka et al., 2016] Hiraoka, T., Neubig, G., Sakti, S., Toda, T., and Nakamura, S. (2016). Learning cooperative persuasive dialogue policies using framing. *Speech Communication*, 84:83 – 96.
- [Hirayama et al., 2010] Hirayama, T., Jean-Baptiste, D., Kawashima, H., and Matsuyama, T. (2010). Estimates of user interest using timing structures between proactive content-display updates and eye movements. *IEICE Transactions on Information & Systems*, E-93D(6):1470–1478.
- [Hofmann, 1999] Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of Uncertainty in Artificial Intelligence, UAI’99*, pages 289–296.
- [Hofmann, 2004] Hofmann, T. (2004). Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, (1):89–115.

- [Ishii et al., 2013] Ishii, R., Nakano, Y. I., and Nishida, T. (2013). Gaze awareness in conversational agents: Estimating a user's conversational engagement from eye gaze. *ACM Transactions on Interactive Intelligent Systems*, 3(2):11:1–11:25.
- [Ishikawa et al., 2012] Ishikawa, E., Yonetani, R., Kawashima, H., Hirayama, T., and Matsuyama, T. (2012). Semantic interpretation of eye movements using designed structures of displayed contents. In *Proceedings of the 6th Workshop on Eye Gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction, Gaze-In'12*.
- [Itti et al., 1998] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.
- [Iwata et al., 2009] Iwata, T., Watanabe, S., Yamada, T., and Ueda, N. (2009). Topic tracking model for analyzing consumer purchase behavior. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence, IJCAI'09*, pages 1427–1432.
- [Jacob and Karn, 2003] Jacob, R. J. K. and Karn, K. S. (2003). Commentary on Section 4. 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*, pages 573–605. Elsevier Science.
- [Jawaheer et al., 2014] Jawaheer, G., Weller, P., and Kostkova, P. (2014). Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems*, 4(2):8:1–8:26.
- [Jayagopi et al., 2012] Jayagopi, D., Sanchez-Cortes, D., Otsuka, K., Yamato, J., and Gatica-Perez, D. (2012). Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proceedings of the 14th ACM international conference on Multimodal interaction, ICMI '12*, pages 433–440.
- [Jin et al., 2004] Jin, X., Zhou, Y., and Mobasher, B. (2004). Web usage mining based on probabilistic latent semantic analysis. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, pages 197–205.

- [Judd et al., 2009] Judd, T., Ehinger, K., Durand, F., and Torralba, A. (2009). Learning to predict where humans look. In *Proceedings of 2009 IEEE 12th International Conference on Computer Vision*, pages 2016–2113.
- [Kaner et al., 2011] Kaner, S., Lind, L., Toldi, C., Fisk, S., Berger, D., and Doyle, M. (2011). *Facilitator’s Guide to Participatory Decision-Making*. Wiley.
- [Kübler et al., 2014] Kübler, T. C., Kasneci, E., and Rosenstiel, W. (2014). Submatch: scanpath similarity in dynamic scenes based on subsequence frequencies. In *Eye Tracking Research and Applications, ETRA ’14*, pages 319–322.
- [Liu and Lane, 2017] Liu, B. and Lane, I. (2017). An end-to-end trainable neural network model with belief tracking for task-oriented dialog. In *Proceedings of Interspeech 2017*.
- [Lowe et al., 2017] Lowe, R. T., Pow, N., Serban, I. V., Charlin, L., Liu, C.-W., and Pineau, J. (2017). Training end-to-end dialogue systems with the ubuntu dialogue corpus. *Dialogue & Discourse*, 8(1):31–65.
- [Mahmood and Ricci, 2009] Mahmood, T. and Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, HT ’09*, pages 73–82.
- [McAllister et al., 1979] McAllister, D. W., Mitchell, T. R., and Beach, L. R. (1979). The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance*, 24(2):228 – 244.
- [Mei et al., 2007] Mei, Q., Shen, X., and Zhai, C. (2007). Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’07*, pages 490–499.
- [Milliman and Decker, 1990] Milliman, R. E. and Decker, P. J. (1990). The Use of Post-Purchase Communication to Reduce Dissonance and Improve Direct Marketing Effectiveness. *The Journal of Business Communication*, 27(2):159–170.
- [Minka, 2000] Minka, T. P. (2000). Estimating a Dirichlet distribution. Technical report.

- [Misu et al., 2011] Misu, T., Sugiura, K., Kawahara, T., Ohtake, K., Hori, C., Kashiooka, H., Kawai, H., and Nakamura, S. (2011). Modeling spoken decision support dialogue and optimization of its dialogue strategy. *ACM Transactions on Speech and Language Processing*, 7(3):10:1–10:18.
- [Nakano and Ishii, 2010] Nakano, Y. I. and Ishii, R. (2010). Estimating user’s engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the 15th International Conference on Intelligent User Interfaces, IUI ’10*, pages 139–148.
- [Orquin and Loose, 2013] Orquin, J. L. and Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, 144(1):190 – 206.
- [Pasupa et al., 2009] Pasupa, K., Saunders, C., Szedmak, S., Klami, A., Kaski, S., and Gunn, S. (2009). Learning to rank images from eye movements. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 2009–2016.
- [Peter and Tarpey, 1975] Peter, J. P. and Tarpey, L. X. (1975). A comparative analysis of three consumer decision strategies. *Journal of Consumer Research*, 2(1):29–37.
- [Pfeiffer et al., 2015] Pfeiffer, J., Duzevik, D., Rothlauf, F., Bonabeau, E., and Yamamoto, K. (2015). An optimized design of choice experiments: A new approach for studying decision behavior in choice task experiments. *Journal of Behavioral Decision Making*, 28(3):262–280.
- [Pohekar and Ramachandran, 2004] Pohekar, S. and Ramachandran, M. (2004). Application of multi-criteria decision making to sustainable energy planning—a review. *Renewable and Sustainable Energy Reviews*, 8(4):365 – 381.
- [Polifroni and Walker, 2008] Polifroni, J. and Walker, M. A. (2008). Intensional summaries as cooperative responses in dialogue: Automation and evaluation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- [Puolamäki et al., 2008] Puolamäki, K., Ajanki, A., and Kaski, S. (2008). Learning to learn implicit queries from gaze patterns. In *Proceedings of the international conference on Machine learning*, pages 760–767.

- [Rayner, 1998] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research.
- [Reynolds and Gutman, 1988] Reynolds, T. J. and Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, 28(1):11–31.
- [Rotgans and Schmidt, 2011] Rotgans, J. I. and Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction*, 21(1):58 – 67.
- [Russo and Leclerc, 1994] Russo, J. E. and Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, 21(2):274–290.
- [Saaty, 1980] Saaty, T. L. (1980). *The Analytic Hierarchy Process, Planning, Priority Setting, Resource Allocation*. McGraw-Hill.
- [Saaty, 2008] Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1):83.
- [Schaffer et al., 2016] Schaffer, E., Kawashima, H., and Matsuyama, T. (2016). A probabilistic approach for eye-tracking based process tracing in catalog browsing. *Journal of Eye Movement Research*, 9(7).
- [Shi et al., 2013] Shi, S. W., Wedel, M., and Pieters, F. G. M. R. (2013). Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science*, 59(5):1009–1026.
- [Shimojo et al., 2003] Shimojo, S., Simion, C., Shimojo, E., and Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12):1317–1322.
- [Steichen et al., 2014] Steichen, B., Conati, C., and Carenini, G. (2014). Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data. *ACM Transactions on Interactive Intelligent Systems*, 4(2):11:1—11:29.
- [Sugano et al., 2013] Sugano, Y., Kasai, H., Ogaki, K., and Sato, Y. (2013). Image Preference Estimation from Eye Movements with A Data-driven Approach. In

Proceedings of the International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction.

- [Tversky, 1972] Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79:281–299.
- [Walker et al., 2004] Walker, M. A., Whittaker, S. J., Stent, A., Maloor, P., Moore, J., Johnston, M., and Vasireddy, G. (2004). Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science*, 28(5):811 – 840.
- [Wentzel and Wigfield, 2009] Wentzel, K. and Wigfield, A. (2009). *Handbook of motivation at school*. Educational psychology handbook series. Routledge.
- [White et al., 2006] White, R. W., Kules, B., Drucker, S. M., and Schraefel, M. (2006). Supporting exploratory search. *Commun. ACM*, 49(4):36–39.
- [Yarbus, 1967] Yarbus, A. L. (1967). Eye movements and vision. *Plenum*.
- [Yonetani et al., 2012] Yonetani, R., Kawashima, H., and Matsuyama, T. (2012). Multi-mode saliency dynamics model for analyzing gaze and attention. In *Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA '12*, pages 115–122.
- [Yonetani et al., 2013] Yonetani, R., Kawashima, H., and Matsuyama, T. (2013). Predicting where we look from spatiotemporal gaps. In *2013 International Conference on Multimodal Interaction, ICMI '13*, pages 421–428.
- [Yoshino and Kawahara, 2015] Yoshino, K. and Kawahara, T. (2015). Conversational system for information navigation based on POMDP with user focus tracking. *Computer Speech & Language*, 34(1):275 – 291.
- [Yoshitaka et al., 2007] Yoshitaka, A., Wakiyama, K., and Hirashima, T. (2007). Recommendation of visual information by gaze-based implicit preference acquisition. In *Proceedings of Advances in Multimedia Modeling: 13th International Multimedia Modeling Conference, MMM 2007*, pages 126–137.