

Semantic Interpretation of Eye Movements Using Author-designed Structure of Visual Content

Erina Ishikawa Schaffer

Abstract

Human eye movements are a powerful descriptor that can help researchers understand how people obtain and process information on computer displays. Existing studies on the analysis of eye tracking data have revealed some of the relationships between human eye movements and their underlying mental states (e.g., interests or intentions). However, understanding content browsing behavior in uncontrolled environments is still a challenge due to the complexity of eye movement patterns. Part of the issue is that human gaze is affected not only by browsers' states but also by the physical and semantic aspects of visual content. Moreover, same eye movement patterns can have different meanings depending on context in content browsing, which may change over time.

First, this study proposes a novel gaze analysis framework to understand the meaning of browsing behavior, which leverages information about the content creator's intentions. Visual content such as web pages, digital articles and catalogs are comprised of structures intentionally designed by content creators, which we refer to as *designed structure*. This study focuses on two design factors of designed structure: spatial structure of content elements (content layout), and their semantic relationships such as "being in a group". The framework was evaluated with an experiment involving twelve participants, wherein the participant's state was estimated from their browsing behavior. The results from the experiment show that the use of designed structure improves estimation accuracy of user states compared to other baseline methods.

Second, this study proposes a hierarchical content browsing model to manage diversity of eye movement patterns. In content browsing situations, human browsers may have different browsing contexts (which we refer to as *browsing states*) and context can change over time. To reveal the characteristics and temporal changes of browsing states in content browsing behavior, this study proposes a probabilistic model based on hidden semi-Markov model (HSMM) whose hidden states correspond to the browsing states. The proposed browsing state model is

trained and evaluated using gaze data that were collected from eight participants in an experimentally controlled catalog browsing situation. We analyze the estimated browsing states and demonstrate that they can contribute to improve the performance of browser interest estimation.

Finally, we adapt the proposed content browsing model to the analysis of choice behavior in digital catalog browsing. Findings from existing studies suggest that the consumer choice process consists of a few different browsing states such as *screening* and *evaluation*. The proposed approach consists of two steps. In the first step, we define and identify the browsing states based on how often a decision maker looks at the selected item (the item finally selected by a decision maker), which acts as a clue into the process of choice behavior. Our definition of browsing states is evaluated through eye tracking data to confirm that it can split the decision process into meaningful browsing states. In the second step, we train our browsing-state model using eye tracking data and corresponding identified browsing states via a supervised manner. Once the browsing state model is learned, we can estimate browsing states of newly observed eye tracking data without the information of the selected item. The latter method is evaluated by measuring how it can replicate the results by the former identification method that uses information of the selected item.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Takashi Matsuyama for the continuous support of my Ph.D study and related research. His vast knowledge, vision, and strong will inspired me through both good and bad times throughout the last six years of my Ph.D study since I was an undergraduate student. I have learned many priceless things from him on both research as well as on my career. I could not have imagined having a better supervisor for my Ph.D study.

I also would like to thank the rest of my thesis committee: Assoc. Prof. Hiroaki Kawashima for his consecutive help and encouragement on my Ph.D study as well as for being a good role model as a young motivated researcher, and Prof. Takatsune Kumada for his insightful comments but also for the critical question which incited me to widen my research from various perspectives.

My sincere thanks go to all the old/current members of Matsuyama Laboratory at Kyoto University for the stimulating discussions, the relaxing coffee breaks, the encouragements at sleepless nights before deadlines, and all the fun we have had in the last six years. I also would like to thank my friends in Santa Barbara for having me during my stay as a visiting researcher at University of California, Santa Barbara, as well as for their warm welcome every time I visited.

I would like to thank my family: my mother, Hideko Ishikawa, and my parents-in-law, William J. Schaffer and Virginia M. Schaffer, for always supporting me spiritually throughout my thesis writing and daily life in general. Last but not the least, I would like to show my special appreciation to my beloved husband, James A. Schaffer, who has been always on my side as a reliable partner, a good counselor, and a best friend.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Background	2
1.2.1	Mental-State Estimation from Eye Movements	3
1.2.2	Application of Estimating Mental States of Browsers	6
1.3	Focus of This Study	7
1.3.1	Basic Assumptions of Content Browsing Situations	7
1.3.2	Issues in the Analysis of Content Browsing Behavior	8
1.4	Contributions	9
1.4.1	A Framework to Encode Eye Movements based on De- signed Structure of Visual Content	10
1.4.2	A Hierarchical Content Browsing Model to Manage Diver- sity of Eye Movement Patterns	10
1.5	Overview of the Thesis	11
2	Hierarchical Modeling of Content Browsing Behavior	15
2.1	Modeling of Content Browsing Behavior	15
2.1.1	The Generative Process of Content Browsing Behavior	16
2.1.2	A Hierarchical Content Browsing Model	16
2.2	Analysis of Content Browsing Behavior	18
2.2.1	Analysis of Eye Movements	18
2.2.2	Identification of Gaze Targets	19
2.2.3	Mining Gaze Actions	19
2.2.4	Modeling Browsing States	20
3	Author-designed Structure of Visual Content	23
3.1	Effects of Content Design on Human Gaze	23
3.2	Definition of Designed Structure	24

3.2.1	Intention-level Designed Structure	24
3.2.2	Appearance-level Designed Structure	26
3.3	Description of Designed Structure	26
3.4	Examples	29
3.5	Discussion	30
4	Encoding Eye Movements into Gaze Actions Using Designed Structure	33
4.1	Introduction	33
4.2	Proposed method	35
4.2.1	Labeling Gaze Transitions Using Designed Structure	35
4.2.2	Analysis of Labeled Gaze Transitions	37
4.3	Experiment	38
4.3.1	Experimental Settings	38
4.3.2	Comparative Eye Movement Features	40
4.3.3	Results	41
4.4	Discussion	43
5	Unsupervised Estimation of Content Browsing States	49
5.1	Introduction	49
5.2	Modeling and Estimating of Content Browsing States Using Hidden Semi-Markov Model	50
5.2.1	Sequence of Gaze Actions	50
5.2.2	Hidden Semi-Markov Model	51
5.2.3	Learning Model Parameters	52
5.3	Experiment	52
5.3.1	Experimental Setting	54
5.3.2	Examples of Collected Gaze Data	55
5.3.3	Results	57
5.4	Discussion	63
6	Model Adaptation for Choice Behavior Analysis	65
6.1	Introduction	65
6.2	Choice Behavior Analysis	67
6.3	Defining Browsing States in Choice Behavior	70
6.3.1	Basic Assumptions of Browsing States	70
6.3.2	The Proposed Choice Behavior Model	71
6.3.3	Evaluation	74

6.4	Supervised Estimation of Browsing States	79
6.4.1	The Overview of The Proposed Estimation Method	80
6.4.2	Modeling Probabilistic Relations of Gaze Actions and Browsing States	81
6.4.3	Evaluation	82
6.5	Discussion	85
7	Conclusion	89
7.1	Summary	89
7.2	Future Work	90
7.2.1	Limitation and Extension of the Proposed Framework . . .	90
7.2.2	Applications	91

List of Figures

1.1	Entities of content browsing situation.	2
1.2	Catalog browsing situation: an example of content browsing situations of interest in this study.	9
1.3	Overview of the thesis.	11
2.1	Generative process of eye movements in goal-directed content browsing.	17
2.2	Overview of the hierarchical content browsing model.	21
3.1	Entities of digital catalogs.	25
3.2	Graph-based representation of content structure.	28
3.3	Examples of catalog layouts.	29
3.4	The intention-level designed structure of the example catalogs. Edges between dotted frames indicate that every node in the frame is connected to every node in the other frame.	30
4.1	Labeling gaze transitions using designed structure.	35
4.2	Encoding of eye movements. (a) Edges between dotted frames indicate that every node in the frame is connected to every node in the other frame. (b) Six patterns of tri-gram paths in the graph G_{IL} . The two bottom features correspond to comparison patterns in which the first and the last dwell of the tri-gram is on the same item.	37
4.3	The experimental environment.	39
4.4	The confusion matrices of the classification of the objectives. The horizontal axis indicates classified objectives and the vertical axis indicates actual objectives. Each number in the matrices indicates the number of gaze sequences.	42

4.5	Out of bag feature importance with category based layout. The feature IDs of horizontal axis correspond to feature ID listed in Table 4.1. The color of bars indicates the type of features.	43
4.6	Out of bag feature importance with order based layout. The feature IDs of horizontal axis correspond to feature ID listed in Table 4.1. The color of bars indicates the type of features.	44
5.1	Overview of the proposed method.	51
5.2	Experimental environment. A participant is browsing a digital catalog on a display. Gaze data of the participant are recorded by the eye tracker installed below the display.	53
5.3	Layout of digital catalogs. Items with a same color hue are in a same group, and each group of items are divided by a gray frame. .	53
5.4	Example of dwell sequences. For each sequence, the color in the top row corresponds to the color of regions in Figure 5.3, and the bottom row shows the timings of dwells on the selected item by light gray highlights.	56
5.5	The AIC values of different numbers of states.	59
5.6	The topology of the trained HSMM.	59
5.7	Output probabilities of each state in HSMM. The color of the bars corresponds to the color of gaze features in Figure 4.2 (b).	60
5.8	Probabilities of duration of each state in HSMM.	60
5.9	Duration distribution of each state before the approximation by a Gaussian distribution.	61
5.10	The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 2). The whiskers represent standard deviation.	61
5.11	The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 3). The whiskers represent standard deviation.	62
5.12	An example timeline of browsing states estimated by the trained HSMM and the trained HMM for the gaze region sequence of the 4th trial of the 8th participants (refer to Figure 5.4).	62
6.1	Segments of browsing states.	73
6.2	The estimated transition probabilities of the browsing states. . . .	75

6.3	The output probabilities of the start state, S_{st} , estimated using eye tracking data of all participants.	75
6.4	The output probabilities of the exploration state (S_{exp}) and the evaluation state, S_{eva} , estimated using eye tracking data of all participants. The bars represent the frequent distribution of segment lengths.	76
6.5	Examples of estimated browsing state sequences. For each sequence, the color in the top row corresponds to the color of regions in Figure 5.3. In the second row, the timings of the dwells on the selected item are shown by light gray highlights. From the third row to the bottom row, the blue and red intervals correspond to the exploration state and the evaluation state, respectively.	77
6.6	The proportion of dwells in each browsing state. The whiskers represent standard deviation.	78
6.7	The proportion of dwells on selected items, non-selected candidates, and the other items when the number of candidates is 2. The whiskers represent standard deviation.	79
6.8	The proportion of dwells on selected items, non-selected candidates, and other items when the number of candidates is 3. The whiskers represent standard deviation.	80
6.9	The output probabilities of each browsing state. The color of the bars corresponds to the color of gaze actions in Figure 4.2 (b). . . .	82
6.10	Example sequences of estimated browsing states. The blue and red intervals correspond to the exploration state, S_{exp} , and the evaluation state, S_{eva} , respectively. The gray intervals correspond to the start state, S_{st}	84
6.11	The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 2). The whiskers represent standard deviation.	85
6.12	The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 3). The whiskers represent standard deviation.	86

List of Tables

4.1	Features used in the experiment. The left column shows the name of features. The second and third column from the left shows feature ID with category-based layout (CL ID) and feature ID with order-based layout (OL ID), respectively. The right column shows the particular labels/relations/values used in the experiment. . . .	46
4.2	The estimation accuracies (precisions) and F measures of the three objectives: input (IN), decision (DE) and free-viewing (FV). The 6 rows from the top show the results with the category-based layout, and the next 6 rows show the results with the order based layout. The first row shows the results obtained using only motion features of eye movements (Gaze). The following four rows show the accuracy obtained combining motion features of eye movements with item-semantic features (Sem), appearance-level (AL) design gaze actions, intention-level (IL) design gaze actions and two content-design-oriented features (AL and IL design gaze actions), respectively. The 6th row shows the accuracy obtained combining motion features of eye movements with all content features: item-semantic features, AL design gaze actions and IL design gaze actions.	47
5.1	The attributes and attribute values used in the experiments.	54
5.2	The number of recorded gaze sequences with each task complexity.	55
5.3	The estimation of the selected item.	63
5.4	The estimation of the set of considered items (selected items and candidates).	63

6.1	The consistency with the ground truth. The top row indicates the consistency when we consider that the exploration state S_{exp} and the start state S_{st} are the same state. The bottom row indicates the consistency when we discriminate the two states.	83
-----	---	----

Chapter 1

Introduction

1.1 Motivation

Eye movements are powerful descriptors for how people obtain and process information on computer displays. Compared to other process tracing methods, such as clickstream or think-aloud protocols, eye movements have the distinct advantage that they can be recorded with higher sampling rates and with fewer cognitive impacts [Glaholt and Reingold, 2011]. Due to the recent improvements in eye tracking techniques, researchers now have easier access to high-fidelity eye tracking data of human browsers. Eye tracking analysis has contributed to various research fields including psychology [Yarbus, 1967], human computer interaction (HCI) [Jacob and Karn, 2003] and marketing research [Wedel and Pieters, 2011].

Yarbus et al. pioneered eye tracking analysis when human browsers are exploring complex images [Yarbus, 1967, Tatler et al., 2010]. They insisted that browsers' eye movements are strongly influenced by their underlying tasks or cognitive goals. Many follow-up studies have examined additional details of the eye-mind link; moreover, some studies attempted to solve its inverse problem, i.e., estimation of browsers' mental states (e.g., tasks) from their eye tracking data [Greene et al., 2012, Borji and Itti, 2014]. Although not all studies have achieved the eye tracking based task estimation successfully [Greene et al., 2012], Borji et al. confirmed that it is possible to estimate human browsers' tasks significantly above chance by combining motion features of eye movements with image features of visual content [Borji and Itti, 2014]. These studies mainly employed statistical features of saccadic eye movements (e.g., the average value of saccade lengths) for the task estimation. Borji et al. discussed about the possibility of im-

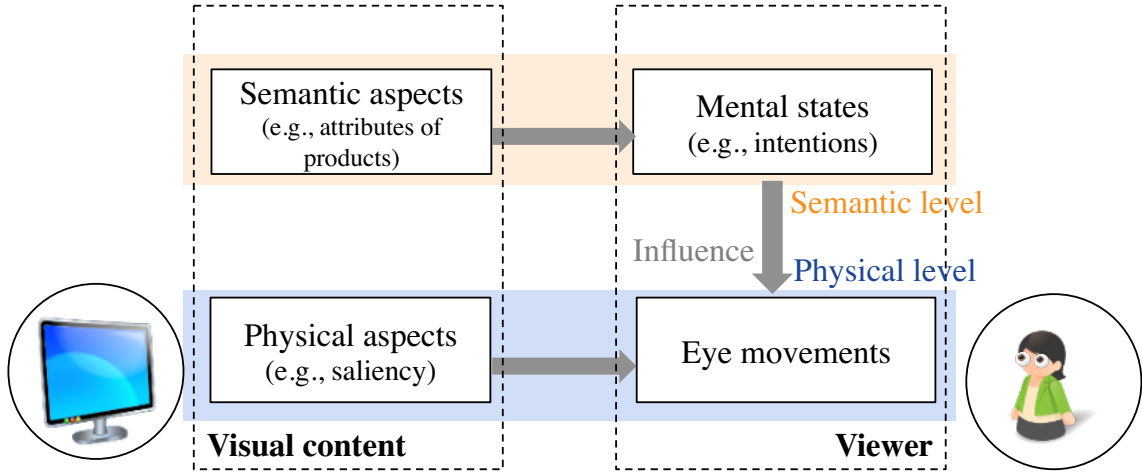


Figure 1.1: Entities of content browsing situation.

proving the estimation accuracy by leveraging more informative features of eye movements [Borji and Itti, 2014]; however, how to extract such informative features from eye tracking data is still an open problem. Since both semantic and physical aspects of visual content affect browsers’ eye movements, not just motion features of eye movements but also information of visual content should be considered to solve the problem.

In this study, we aim to take another step for mental state estimation from eye movements by establishing an eye-tracking-analysis framework for *semantic interpretation of eye movements*. When people are performing a complex task, their content browsing behavior is expected to consist of a series of actions caused by the task and their underlying intentions. For example, when people are making a decision among multiple alternatives, they need to commit acts such as “a binary comparison of alternative A and B” or “elimination of alternative C”. The goal of this study, semantic interpretation of eye movements, is to associate those gaze actions with short-term eye movement patterns. Understanding gaze actions during a human browser is performing a task enables deeper understanding of the underlying process until the browser accomplishes the task beyond mere estimation of which task the browser is performing.

1.2 Background

Eye movements have long been an important cue for underlying mental states of human browsers in various situations including human-agent conversation, in-

formation search, decision making, and so on. The term *mental states of browsers* in this study indicates human internal states that influence content browsing behavior. This section briefly summarizes representative works on mental state estimation from eye movements and their applications.

1.2.1 Mental-State Estimation from Eye Movements

Estimating browsers' mental states from eye movements has been of interest especially in the last decade due to easy access to eye tracking data. Probabilistic approach employing machine learning techniques is commonly taken to manage the complex eye-mind link in modern eye tracking studies. For this, eye tracking data are first encoded into statistical features, which we call *eye-movement features*, and then probabilistic classification method is applied to the features. In this section, we categorize existing studies on mental state estimation using eye movements based on two aspects: (A) how to characterize eye-movement features, and (B) types of classification methods.

(A) Eye-Movement Features

Basic representation of eye tracking data is to employ motion features of eye movements, which is simple and fully content-independent. Meanwhile, some studies employed information of visual content to characterize eye movements since eye movements are influenced by both physical and semantic structure of visual content (Figure 1.1). The following paragraphs briefly introduce both motion features of eye movements and eye-movement features characterized using content information which are used in existing studies for mental-state estimation.

(A-1) Motion Features of Eye Movements

Below we will specify some technical terms to represent motions of eye movements. This study focuses on static visual content which does not include any dynamics. When a human is browsing static visual content, the motion of eye movements can be categorized into two types: fixation and saccade.

Fixation A fixation or visual fixation occurs when gaze fixates on a single location from approximately 100 msec to more than 1 sec (depending on the situation of content browsing).

Saccade A saccade is a high-speed eye movement that occurs between fixations. No information can be obtained as perception (saccadic suppression); thus, almost all information is acquired during fixations. Each saccade requires approximately 30 msec for 5° of visual angle and 2 msec longer for each additional degree.

The motion features of eye movements here indicate the characteristics of the above two types of eye movements. Existing studies often use multiple features related to fixations and saccades; for example, fixation positions/durations/numbers and saccade lengths/directions/durations. Statistical approaches using motion features of eye movements are considered to be a robust and content-independent method for gaze analysis and they are utilized in many different situations [Sugano et al., 2014, Pasupa et al., 2009, Bednarik et al., 2012, Eivazi and Bednarik, 2011]. For example, Sugano et al. classify motion features of eye movements using a Random Forest algorithm to estimate browsers' preference towards pictures [Sugano et al., 2014]. In [Bednarik et al., 2012], motion features of eye movements are learned by using Support Vector Machines to discriminate non-intentional and intentional eye movements.

Above mentioned studies mainly consider statistics of motion features extracted from observed gaze sequence (e.g., mean or variance). However, the statistics of motion features alone discard temporal information in the eye tracking data, which can be an important clue to understand content browsing behavior. Some repetitive eye movement patterns that consist of multiple saccades and fixations may be more informative features that represent important aspects of browsing behavior, e.g., "scanning the scene widely" or "examining in a small area". Bulling et al. proposed a method to incorporate such repetitive eye movement patterns for eye tracking based activity recognition [Bulling et al., 2011]. In their method, saccades are first encoded based on length and direction into 24 discrete characters, then tri-gram analysis is employed to the encoded saccades sequence. Some other studies applied their eye movement features to activity recognition in different domain such as cooking behavior recognition [Inoue et al., 2016].

(A-2) Incorporating Physical and Semantic Aspects of Visual Content

The attempt to incorporate physical aspects of visual content to characterize eye movement features has been done by some studies [Sattar et al., 2015, Borji and Itti, 2014]. Sattar et al. employed statistical image features (local binary

patterns) around fixations to predict search targets [Sattar et al., 2015]. Moreover, based on findings in the field of visual psychology, human eye movements are known to be more attracted by an area if it is physically distinguished from its surroundings, and many studies have proposed a model to predict the degree to which visual areas attract attention, referred to as *saliency*. Saliency is also employed to characterize eye movement features; for example, Borji et al. incorporated saliency values of fixations as Normalized Scanpath Saliency (NSS) values to estimate tasks [Borji and Itti, 2014]. NSS is a measurement to assess saliency prediction models proposed by Peters et al. [Peters et al., 2005], and it can be calculated as the average of the response values at all fixations in a normalized saliency map.

Incorporating semantic aspects of visual content is based on more heuristic approaches, where semantic labels are manually annotated to areas-of-interest (AOIs). For example, Nakano et al. used semantic information of visual content to estimate browsers' engagement in conversation with displayed agents [Nakano and Ishii, 2010]. Semantic labels such as "the agent's head" and "the object that the agent is explaining" are annotated with regions on the screen, and N-gram analysis is applied to a sequence of the semantic labels on AOIs. This approach is considered to be a simple and effective way to introduce semantic information of visual content into eye tracking analysis. However, it can only deal with analyst-specified domains/situations since it becomes difficult to determine appropriate labels for object regions as visual content becomes complex and diverse. For example, as the amount of objects in visual content increases, the number of semantic labels will become unmanageable.

(B) Probabilistic Models for Mental-State Estimation

As mentioned above, eye movements in an uncontrolled situation can be very complex because they are influenced by both browsers' mental states and physical and semantic structure of visual content. Thus, existing studies employ probabilistic approaches to manage the complexity of eye movements. This section categorizes those probabilistic approaches into two types based on the employed learning method: (1) when target mental states are known or well-defined and we can obtain eye tracking data with associated ground truth, mental state estimation can be done in a straightforward supervised manner; (2) when target mental states are unknown and/or when they may change temporally, an unsupervised approach is necessary.

(B-1) Supervised Methods for Estimating Static, Pre-defined and Knowable Mental States

In some situations, mental states such as tasks and preference of browsers are not expected to change during the observation of eye movements. For these situations, it is comparatively easy to obtain ground truth of mental states when eye movements are being observed simply using think-aloud method or controlling task/conditions. Once we obtain complete data (a set of eye tracking data and its ground truth of associated mental states), we can train supervised classification model with the data.

Both linear classification methods [Pfeiffer et al., 2014, Greene et al., 2012] and non-linear classification methods such as ensemble learning [Borji and Itti, 2014, Sugano et al., 2014] are used in existing studies.

(B-2) Unsupervised Methods for Estimating Dynamic, Undefined and Unknowable Mental States

Unsupervised methods are used when target mental states can change during observation of eye movements or when we do not have any prior knowledge about mental states we want to estimate. Brandherm et al. attempted to estimate the interest of browsers in a virtual sales scenario based on dynamic Bayesian networks [Brandherm et al., 2007]. State space models such as a hidden Markov model is also used to model transitions among information processing states in information search and in decision making [Simola et al., 2008, Shi et al., 2013].

1.2.2 Application of Estimating Mental States of Browsers

Understanding mental states in browsing visual content such as digital articles, web pages, and catalogs is of great interest for various applications. For example, recommendation/information systems can be improved by understanding users' mental states such as interest [Hirayama et al., 2011, Shimonishi et al., 2016] since it enables the systems to provide appropriate information at right time. Moreover, supporting systems for students in watching video lectures [Rodrigue et al., 2015] and human-agent interaction [Nakano and Ishii, 2010] can be improved by understanding content browsing states of users.

1.3 Focus of This Study

This section is to clarify focal problems that this study tackles. First, we describe the basic assumptions of content browsing situations. Then, we present two main issues to achieve semantic interpretation of eye movements in the assumed content browsing situation.

1.3.1 Basic Assumptions of Content Browsing Situations

Assume the content browsing situation where a human browser is viewing a screen and the browser's eye movements are recorded as a sequence of gaze points. This study focuses on content browsing situations where (1) visual content contains some artificial structure designed by content creators and (2) a person is browsing the visual content with some purpose. The content browsing situations of interest include online shopping, information search on the Web, reading magazines, etc. The following paragraphs describe each condition particularly.

Visual content with author-designed structure Visual content in this study contains multiple objects described in various media (e.g., text and images). For simplicity, this study assumes the region boundaries of the objects are obvious so that AOIs can be easily segmented. We also assume the content to have designer/author-imposed structure; that is, the objects consist of items are placed based on the content creator's intention. Visual content without any artificial structure, e.g., natural images, is not the interest of this study. Finally, we only consider static visual content that contains no dynamic objects; that is, video content is not of interest in this study.

In summary, this study assumes that visual content satisfies the following requirements.

- Visual content contains multiple objects with obvious region boundaries, and the regions do not overlap each other.
- The placement of object regions reflects the content creators' intentions.
- Visual content does not contain dynamic objects (static visual content)

Goal-directed content browsing behavior As mentioned in Section 1.1, the goal of this study, semantic interpretation of eye movements, indicates to associate short-term eye movement patterns with gaze actions that reflect browsers'

mental states. Since it is hard to define and interpret gaze actions in unintentional/unconscious content browsing situations where browsers do not pay attention to the visual content, we only consider goal-directed content browsing situations where browsers have a specific goal/task to accomplish and they browser the visual content intentionally. Although detecting such low attentive states to visual content is one of the interesting topics in eye tracking analysis [Yonetani et al., 2012], such unintentional/unconscious browsing behavior is not of interest in this study.

As an example of content browsing situations that satisfy above requirements, this study mainly studies multi-alternative choice behavior in digital catalog browsing. As shown in Figure 1.2, a browser selects an item from a digital catalog which contain information about multiple items. In the catalog browsing situation, the placement of items of the catalog reflects content creators intentions such as groupings of items, and the browsers make a decision after evaluating alternatives as much as they need.

1.3.2 Issues in the Analysis of Content Browsing Behavior

This section summarizes general issues to achieve detailed understanding of content browsing behavior. As reviewed in Section 1.2.1, many researchers aimed to model relations between people’s eye movements and mental states, and proposed a method to estimate mental states from eye movements. In spite of those works, it is still challenging to establish a versatile method of mental state estimation that can be applied to complex real-world environments. This is because of (1) the diversity of visual content, and (2) the uncertainty in the meaning of eye movement patterns.

In particular, this study seeks to address the following issues to achieve semantic interpretation of eye movements.

1. How can we employ diverse visual content to encode eye movements? As described in Section 1.2.1, incorporating information of visual content is important to understand eye movement patterns since eye movements are influenced not only by mental states but also by visual content. In existing studies, semantic labels which were annotated to AOIs on-screen were used to consider semantic aspects of visual content to understand browsing behavior. However, this approach can only deal with pre-specified domains and situations due to difficulty in determining the correct labels for AOIs as visual content becomes more com-

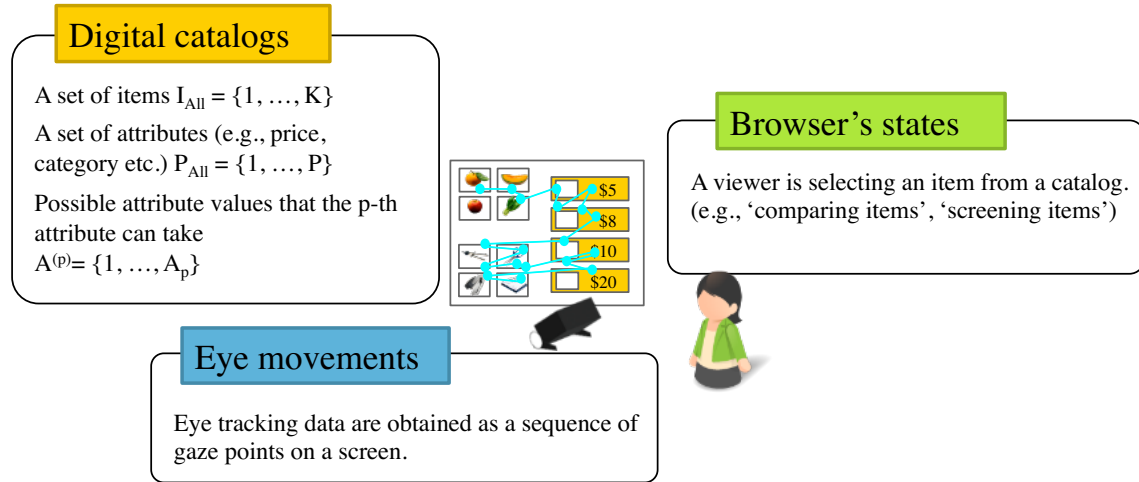


Figure 1.2: Catalog browsing situation: an example of content browsing situations of interest in this study.

plex. For example, the number of semantic labels will increase as the number of AOIs increases.

2. How can we manage uncertainty in the meaning of eye movement patterns?

Eye movement patterns can have different meanings depending on the context of content browsing. Assuming a catalog browsing situation, comparison patterns among a few items in the catalog may be interpreted as “being interested in the items” if it is temporally close to a decision being made, but they would be better interpreted as “screening candidates” if it is at the beginning of choice process. Thus, we need to consider contexts of content browsing to resolve such uncertainties.

1.4 Contributions

The main contributions of this study are two-fold: (1) a novel framework to characterize eye movement patterns using semantic/spatial structure of content design (designed structure), and (2) a hierarchical content browsing model to manage diversity of eye movement patterns.

1.4.1 A Framework to Encode Eye Movements based on Designed Structure of Visual Content

The first contribution of this study is to propose a framework that incorporates spatial and semantic structure of visual content designed by content creators, which we call *designed structure*. For the first issue mentioned in Section 1.3.2, this study employs relationships among the objects instead of annotating semantic labels of content objects. In the proposed framework, browsing behavior is characterized by which part of designed structure attracts the focus and how behavior is influenced by designed structure, e.g., “looking at items in order” or “comparing items in a group”. Introducing designed structure enables a simple representation of visual content, and enables domain independent analysis.

To achieve this, we first model designed structure of visual content by a directed graph. The details of definition and representation of designed structure will be described in Chapter 3. Then, we propose an analysis method of eye movements based on designed structure (see Chapter 4 for details). The proposed framework is evaluated through an eye tracking experiment in a catalog browsing situation by measuring accuracy of mental-state estimation.

1.4.2 A Hierarchical Content Browsing Model to Manage Diversity of Eye Movement Patterns

The second contribution of this study is to propose a hierarchical content browsing model to manage diversity of eye movement patterns. For the first issue mentioned in Section 1.3.2, this study explicitly models context in content browsing, which we call *browsing states*. Differing from preference or tasks of browsers, the browsing states may change during content browsing and are assumed to not be observable. We employ a state space model to represent the browsing states as hidden states, and characterize eye movement patterns by both designed structure and estimated browsing states.

In this study, transitions of browsing states are modeled by a hidden semi-Markov model with several discrete states that can be learned from eye tracking data in a unsupervised manner. The details of the representation of the proposed model and its learning method are described in Chapter 5. Moreover, the model can be learned in a supervised manner when browsing states can be directly defined using prior knowledge in a certain domain. We focus on a catalog browsing

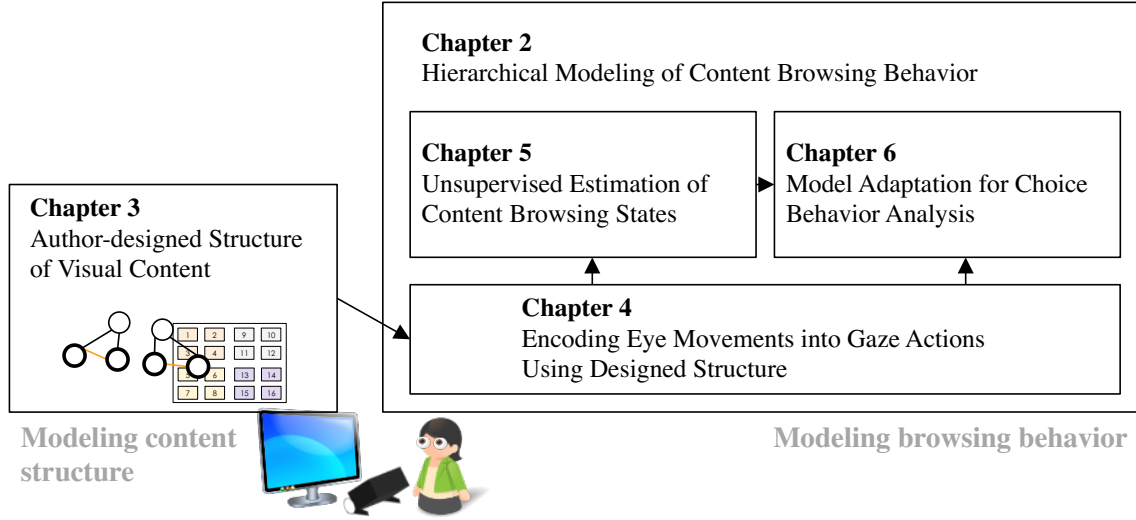


Figure 1.3: Overview of the thesis.

situation and propose an analysis method of eye movements in consumer decision process in Chapter 6. The validity of the proposed model is also evaluated through an eye tracking experiment.

1.5 Overview of the Thesis

This section presents an overview of this thesis. The structure of the following chapters in this thesis is shown in Figure 1.3. Chapter 2 presents an overview of modeling content browsing behavior in this study, and Chapter 3 describes the details of modeling designed structure of visual content. The remaining chapters present the components of the proposed content browsing model in Chapter 2. Using designed structure defined in Chapter 3, eye movements are first encoded into gaze actions in Chapter 4. Then, browsing states are estimated from the sequence of gaze actions in Chapter 5. Finally, Chapter 6 proposes a choice behavior model to understand consumer decision process in a catalog browsing situation based on the proposed content browsing model in Chapter 2.

Chapter 2: Hierarchical Modeling of Content Browsing Behavior

We propose a hierarchical content browsing model to understand goal-directed content browsing behavior. In the proposed model, content browsing behavior is represented with hierarchical structure to manage its diversity. The context of content browsing (i.e., browsing states) is modeled and estimated in the top layer,

and eye movement patterns are interpreted using the estimated browsing states.

In Chapter 2, we first present the basic assumptions of the generative process of content browsing behavior, and the hierarchical content browsing model is proposed based on the assumptions. The definition and details of the layers in the hierarchical content browsing model is also described. Second, we introduce some related studies to obtain eye movement representation in the each layer of the hierarchical content browsing model. Finally, we introduce some existing analysis methods of eye movements to compare with the proposed content browsing model.

Chapter 3: Author-designed Structure of Visual Content

We propose spatio-semantic structure of visual content for eye movement analysis, which we call designed structure. Designed structure is considered at two levels: spatial structure of content elements (appearance-level), and their semantic relations such as “being in the same group” (intention-level). Each level of designed structure represents different aspects of content design.

In Chapter 3, we first introduce some existing studies that investigated the effects of content design on human gaze. Second, we present the definition of intention-level and appearance-level designed structure. Third, we describe the representation of designed structure using a directed graph whose nodes and edges represent objects in visual content and their design relations, respectively. Finally, we present designed structure of two types of catalogs as examples.

Chapter 4: Encoding Eye Movements into Gaze Actions Using Designed Structure

In Chapter 4, we evaluate the framework to encode eye movements into meaningful eye movement patterns, which we call *gaze actions*, using the designed structure defined in Chapter 3. In the proposed method, observed eye tracking sequences are characterized based on semantic and spatial relations in designed structure of visual content.

In the proposed method, eye movement patterns are characterized by which part of designed structure attracts the focus of the browser and how behavior is influenced by designed structure. The proposed method is evaluated with an eye tracking experiment in a catalog browsing situation by measuring accuracy of task estimation compared to other baseline methods to characterize eye move-

ment patterns.

Chapter 5: Unsupervised Estimation of Content Browsing States

In Chapter 5, we introduce a hidden semi-Markov model (HSMM) to represent temporal changes of browsing states. In the proposed gaze model, the states in consumer decision processes are explicitly modeled as hidden states to reveal the characteristics and temporal changes of browsing states in catalog browsing situations. Using the method described in Chapter 4, eye movements are first encoded into a sequence of gaze actions using designed structure. Then, the proposed gaze model is trained and evaluated on real gaze data.

We first present an overview of the browsing model and then explain the training method of the proposed gaze model. To evaluate the proposed model, we collected gaze data from eight participants in an experimentally controlled catalog browsing situation. Using the data, we analyze the estimated browsing states and demonstrate that the states can contribute to improve the performance of browser interest estimation.

Chapter 6: Model Adaptation for Choice Behavior Analysis

In Chapter 6, we adapt the proposed framework proposed in Chapter 5 to the analysis of consumer decision processes. Findings from existing studies suggest that the consumer decision process consists of a few different browsing states such as “screening” and “evaluation”. We first define browsing states based on how often a decision maker looks at the selected item (i.e., the item finally selected by a decision maker) and propose a probabilistic model to identify the browsing states. After confirming that our proposed definition can split decision process into meaningful browsing states through eye tracking data, we train the content browsing model using hidden semi-Markov model by the identified browsing states in a supervised manner.

Chapter 2

Hierarchical Modeling of Content Browsing Behavior

This chapter presents an overview of our framework of modeling content browsing behavior. In the following sections, we first present basic assumptions to model content browsing behavior; and then, we propose a hierarchical content browsing model based on the assumptions. Second, based on the hierarchical content browsing model, we introduce analysis methods of content browsing behavior. More specifically, we describe how to convert observed eye tracking data from the bottom layer to the higher layers in the hierarchical content browsing model by introducing some related studies.

2.1 Modeling of Content Browsing Behavior

Recall that the goal of this study, semantic interpretation of eye movements, is to associate short-term eye movement patterns of browsers with their intended gaze actions. Thus, this study assumes goal-directed content browsing situations, where a browser is trying to accomplish a specific task/goal. To achieve semantic interpretation of eye movements in such situations, we propose a hierarchical content browsing model. The proposed model enables representation of complex and diverse content browsing behavior.

In the following sections, we first present the generative process of eye movements in the goal-directed content browsing situations. Based on the generative process, we model content browsing behavior with a hierarchical structure that consists of four different layers.

2.1.1 The Generative Process of Content Browsing Behavior

Figure 2.1 shows the overview of the generative process of eye movements in the goal-directed content browsing situations. Such situations include information searching, multi-alternative choice, and so on. In the following paragraphs, we take multi-alternative choice as an example for the explanation of the generative process.

1. From a task to subtasks. When a given task is complex enough, a browser is expected to split the task into multiple subtasks to make the task easier to accomplish. For example, when the browser is making a choice among a set of items on a screen, the browser may split the decision making task into a set of smaller tasks such as “screening of items” and “evaluation of strong candidates”.

2. From a subtask to gaze actions. Each subtask can be accomplished by a series of actual actions, which we refer to as *gaze actions*. The gaze actions in this study are defined to be the minimum units of content browsing behavior that involves multiple objects on a screen, e.g., “comparing item A and B”. Moreover, the gaze actions induced by subtasks can differ among the subtasks. For example, when a browser is in the subtask of “evaluation of strong candidates”, gaze actions to compare items (e.g., more frequent gaze transitions among a few items) can occur with higher frequency than in the subtask of “screening of items”.

3. From a gaze action to gaze targets. Gaze actions consist of a series of glances on target objects on a screen. For example, the above mentioned gaze action of “comparing item A and B” consists of alternative glances on item regions of item A and B.

4. From a gaze target to eye movements. Finally, each glance on a target is realized as a set of fixations and saccades.

2.1.2 A Hierarchical Content Browsing Model

Based on the assumption in the previous section, we propose a hierarchical content browsing model. The overview of the hierarchical model is shown in Figure. 2.2. The model consists of four different layers of browsing behavior representation, and each layer represents an important aspect of content browsing

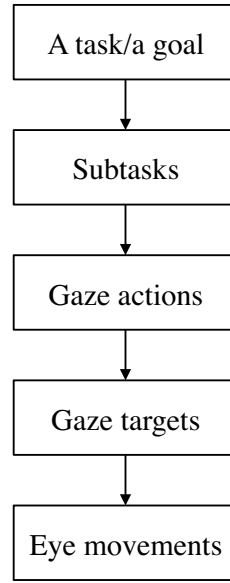


Figure 2.1: Generative process of eye movements in goal-directed content browsing.

behavior, respectively. More specifically, each layer corresponds to each step in the generative process explained in Section 2.1.1.

The following paragraphs present the definition of browsing behavior representation in each layer of the model.

The First Layer: Eye Movements

The first layer (the bottom layer) simply corresponds to *eye movements* in the generative process shown in Figure 2.1. Note that only this layer is observable; that is, we can directly obtain it as the outputs of an eye tracker. As described in Chapter 1, eye movements consist of a few types of motions. For the static content browsing situation, basically, we only need to consider saccades and fixations.

The Second Layer: Unary Gaze Targets

The second layer corresponds to *gaze targets* in the generative process shown in Figure 2.1. This layer represents “which object is looked at”, which is important when modeling interest/preference toward targets on a screen.

The Third Layer: Multi-object Gaze Actions

The third layer corresponds to *gaze actions* in the generative process shown in Figure 2.1. This layer represents “why the objects are being looked at”. However, the

semantic interpretation of gaze actions can vary depending on browsing context. For example, if items in a same group are looked at, it can be interpreted as “the user is interested in the group” or “the user is exploring items along the layout”. This issue was noted in Section 1.3.2 as one of the general issues that this study tackles.

The Fourth Layer: Browsing States

The fourth layer corresponds to *subtasks* in the generative process shown in Figure 2.1. This layer represents “why the gaze actions occurred”, and plays an important role to solve the issue mentioned in the previous paragraph by giving an interpretation to gaze actions.

2.2 Analysis of Content Browsing Behavior

This section introduces related studies for analyzing content browsing behavior based on the hierarchical content browsing model. We introduce some related studies that convert one layer to an upper layer in the hierarchical model. For the analysis phase, we are basically able to observe only eye tracking data (i.e., the first layer in the model). Thus, the analysis here is to convert eye tracking data into content browsing representation in the higher layers of the model.

Since the focus of this study is to obtain gaze actions in the third layer and to model browsing states in the fourth layer, we use simple methods for the analysis in the first layer and the second layer. The details are described below.

2.2.1 Analysis of Eye Movements

Although eye tracking data can be recorded with less effort thanks to modern eye tracking devices, detecting fixations (and saccades) robustly is still challenging in real-world settings due to measurement noise and individual variability. Basically, most studies simply set a threshold on the velocity of transitions of gaze points or on the size of region in which some gaze points stay to identify saccades and fixations. In this study, we simply employ the method that uses a threshold on the velocity of gaze points (Section 4.3.2).

Some studies aim to tackle this problem using different algorithms [Salvucci and Goldberg, 2000, Salvucci and Anderson, 2001,

Li et al., 2016]. For example, Salvucci et al. evaluated three tracing methods that are not only fast but robust to noise and individual variability for automated analysis of saccadic eye movements. In these methods, fixations are modeled with a hidden Markov model (HMM) [Salvucci and Anderson, 2001].

2.2.2 Identification of Gaze Targets

Identifying gaze targets is also challenging in real-world settings due to the measurement noise. In this study, we simply identify gaze targets by referring to positions of gaze points and annotated boundaries of object regions assuming obtained eye tracking data are accurate enough. This can be done because object regions are supposed to have clear boundaries and not overlap with each other in our content browsing situation.

However, in real-world environments, defining object regions and identifying gaze targets can be a chicken-and-egg problem since it is sometimes difficult to annotate appropriate object regions even for human annotators. Some clustering method of gaze points and identification method of gaze targets are proposed. For example, Santella et al. proposed a clustering method using the mean shift algorithm [Santella and DeCarlo, 2004].

2.2.3 Mining Gaze Actions

N-gram is commonly used to represent eye movement patterns for mining meaningful gaze actions [Nakano and Ishii, 2010, Kübler et al., 2014].

In previous studies, semantic annotation labels that were assigned to areas-of-interest (AOIs) on-screen were used to understand the content semantics of gaze behavior. For example, Nakano et al. used tri-gram patterns of a dwell sequence on AOIs with assigned semantic labels such as “the agent’s head” and “the object that the agent is explaining” for estimating the degree of viewers engagement in human-agent conversation [Nakano and Ishii, 2010].

This approach is considered to be a simple and effective way to introduce semantic information of visual content. However, it can only deal with the specified domains/situations since it becomes difficult to determine appropriate labels for object regions as visual content becomes complex and diverse. For example, as the number of the objects contents in the visual content increases, the number of semantic labels will become larger.

2.2.4 Modeling Browsing States

Some previous studies took an approach to model temporal changes of information processing states in search behavior or decision making using state space models. Liechty et al. proposed a HMM to capture two distinct attention states (local or global) while visually exploring complex scenes [Liechty et al., 2003]. In the local search state, viewers focus on specific aspects and details of the scene; meanwhile, in the global attention state, viewers focus on exploring the informative and perceptually salient areas of the scene. Their model is evaluated through an eye tracking experiment of consumers' attention toward print advertisements in magazines.

Simola et al. [Simola et al., 2008] employed a HMM to uncover processing states in text reading. Three HMMs were built using sequences of gaze motion features (e.g., saccade lengths) corresponding to three different information-search tasks: simple word search, finding the answer to a question, and selecting an interesting title from a list. In their model, hidden states are learned using gaze data without any prior knowledge, and are interpreted as *scanning*, *reading*, and *decision*. Their results showed that the tasks can be estimated from a newly observed sequence of eye movements by comparing the likelihood of each model.

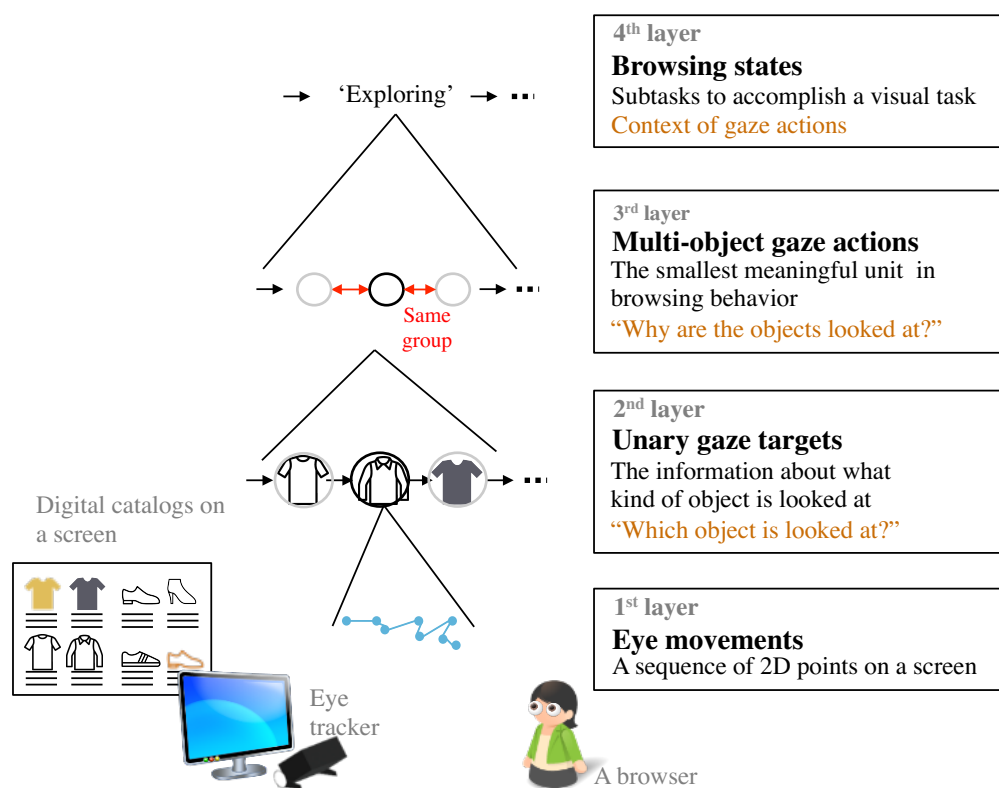


Figure 2.2: Overview of the hierarchical content browsing model.

Chapter 3

Author-designed Structure of Visual Content

In this chapter, we present the definition of designed structure and its representation. First, we introduce some existing studies that investigated the effects of content design on human gaze. Second, we define designed structure of visual content. The designed structure here is considered at two levels: intention-level (IL) and appearance-level (AL). Both levels of designed structure describe different aspects of content design. Third, we describe the representation of designed structure using a directed graph whose nodes and edges represent objects in visual content and their design relations, respectively. Finally, we give a few examples of designed structure assuming a catalog browsing situation as an example of content browsing situations.

The details about how to encode eye movements using the defined designed structure is described in Chapter 4.

3.1 Effects of Content Design on Human Gaze

Eye movements of content browsers are affected by content design such as the layout of object regions. For example, during content browsing with a list layout, browsers tend to look at high ranking objects in a list and pay less attention to objects placed at the bottom. Some existing studies in the field of designing web pages/user interface investigated how placement and organization of object regions on visual content influence human gaze [Chen and Pu, 2011, Pan et al., 2004]. For example, Chen et al. explored differences of content browsing behavior among recommender systems with a simple list layout and an orga-

nization layout where recommended items are grouped into categories and each category is annotated by a title [Chen and Pu, 2011].

Main research interest of the existing eye tracking studies of content design was to employ findings from the analysis to improve content design/usability. To our best knowledge, no study has aimed to exploit information about content design to understand the meaning of browsing behavior.

3.2 Definition of Designed Structure

In this section, the definition of designed structure is introduced in more detail. Designed structure is spatio-semantic structure of visual content intentionally designed by content creators. While humans are browsing visual content, their eye movements are expected to be affected by the designed structure of the page they are viewing. Simultaneously, humans may have their own intentions while content browsing, which naturally affects their eye movements. As described in Chapter 2, the novelty of the proposed framework for encoding eye movements in this study is to introduce the designers' perspective as one of the entities constituting the catalog browsing situation.

Designers compose visual content using various media (e.g., images and text) and decorations (e.g., icons and frames) based on pre-existing design criteria such as "place semantically related items close together". The composition of the resulting content structure is what we call designed structure. For example, in this study, we leverage group relations among objects in visual content as the most basic design criteria for the purpose of understanding browsing behavior. That is, the content structure is represented by a set of objects intentionally grouped by designers. The designed structure in this study is considered at two levels: intention-level (IL) and appearance-level (AL). The former expresses high-level semantic relations among items that reflect designers' intention (e.g., grouping of objects), and the latter expresses low-level appearance information of visual content (e.g., spatial placement of objects). The following paragraphs present details of both intention-level and appearance level designed structure.

3.2.1 Intention-level Designed Structure

Designers usually have prior intentions for visual content and compose the content structure by emphasizing semantic relationships among items, such as "these

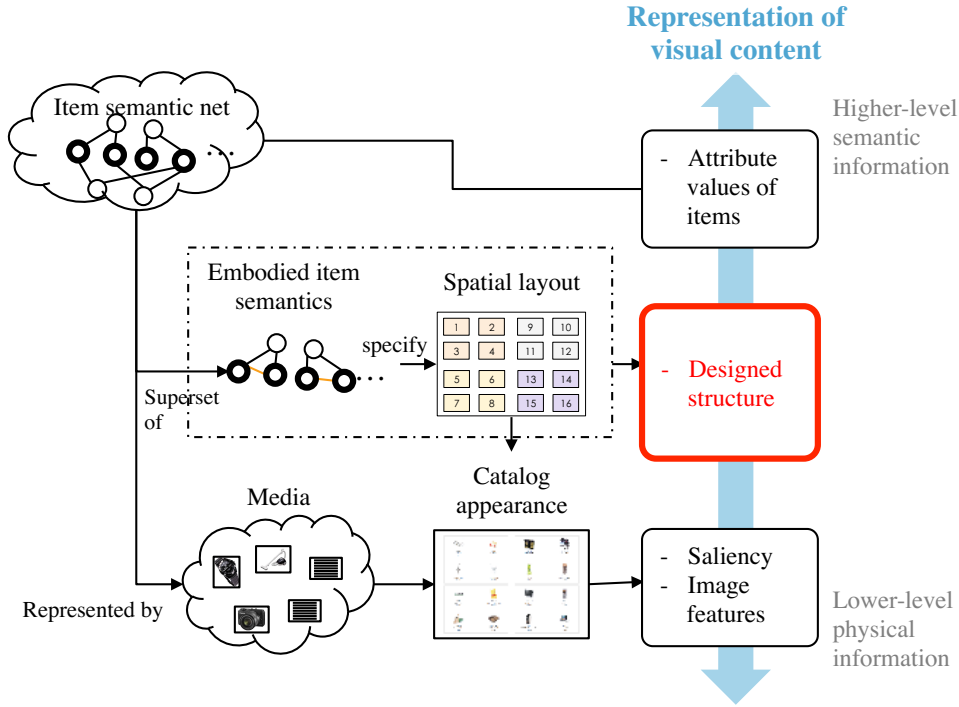


Figure 3.1: Entities of digital catalogs.

items are in the same group”. We call the semantic relationships emphasized by designers *design relations*, and use them to characterize gaze transitions. As the most common design relationships, this study deals with the following:

- **Parallel relation**: Indicates similarity of items, e.g., items in the same category are linked with this relation.
- **Contrast relation**: Indicates the difference of items.
- **Ordinal relation**: Indicates the order of items, e.g., items sorted by their ratings are linked with this relation.

These relations can be translated from the item groups defined by content designers (details will be described in Sec. 3.3). Although other types of relationships can also be considered, such as *part-of* relations and *abstract-of* relations, the following parts of this thesis only consider the above three relations for simplicity.

In this study, we assume the designers’ intention is given information. This assumption is reasonable since the analysts of browsing behavior can access the information of design relations among items in many cases (e.g., the case that the analyst prepares visual content). However, it is not always true that the analysts can access information, such as the case where analysts do not prepare visual

content (for example, analysis of information search behavior on the Web). We discuss a possible method to acquire the intention-level design structure from existing visual content in the following discussion section (Sec. 3.5).

3.2.2 Appearance-level Designed Structure

The information of items is described by various media such as images and text in visual content. The media regions are arranged based on a certain layout. According to the findings from previous gaze studies, the region positions are an important factor that affects browsers' gaze behavior [Goldberg et al., 2002]. In this study, we employ spatial layouts as content appearance information to characterize eye movements. In particular, this study employs spatial relations among object regions such as "far from each other" to interpret gaze transitions among object regions on the screen.

We assume that the intention of content designers and the appearance of content are highly related with each other. That is, content designers decide which types of media to be used and compose the structures of the media regions, layouts, and formats to represent their intents (e.g., items being the same group would be placed close to each other). In other words, designed structure is a realization of content designers' intentions. Additionally, it is the purpose of design conventions and rules to make the content information comprehensible to potential browsers. Design conventions and rules to realize designed structure as actual visual content are discussed in more detail in Sec. 3.5.

3.3 Description of Designed Structure

Here we assume a catalog browsing situation to describe designed structure. Suppose a browser is browsing a digital catalog to select a gift for his/her friend. The digital catalog contains description of several items via various media such as images and text. The catalog consists of three entities: item semantic attributes, intention-level designed structure, and appearance-level designed structure.

Description of Item Semantic Attributes

Items in a digital catalog have various semantic attributes such as brand and category. Semantic attributes are shared by some items in the catalog and shared

attributes can be considered to be important in the analysis of gaze transition behavior. For example, if a browser is comparing items, it is possible that the common attributes of the compared items are of interest. However, as the number of the items increases, the semantic attributes of items can be more diverse. Thus, when the gaze data is limited, it becomes difficult to obtain meaningful interpretation of browsing behavior using item semantic attributes.

Suppose a catalog contains information of a set of items $\mathcal{I}_{All} = \{1, \dots, N\}$ (see in Figure 3.2 (a)). Each item has a set of P attributes $\mathcal{P}_{All} = \{1, \dots, P\}$, where p -th attribute can take a value of A_p possible attribute values $\mathcal{A}^{(p)} = \{a^1, \dots, a^{A_p}\}$. Some attributes can have ordinal relations with each other such as *rating*. In that case, the index of the attribute values correspond to the ordinal relations, that is, if $a > a'$ ($a, a' \in \{1, \dots, A_p\}$), the a -th attribute value is larger than the a' -th attribute value. We introduce a function $f_p : \mathcal{I}_{All} \rightarrow \mathcal{A}^{(p)}$, where $f_p(i)$ indicates the attribute value of the p -th attribute that the i -th item has.

Description of Intention-level Designed Structure

When content designers create a digital catalog, they decide which aspects of content should be emphasized and allocate items based on the certain criteria (intention-level designed structure). For example, if the designer emphasizes a specific attribute (e.g., *category*), all items in the same category are regarded as “in the same group”. To represent this process, we introduce a set of the emphasized attributes as $\mathcal{P}_F \subseteq \mathcal{P}_{All}$. Moreover, the emphasized attributes are categorized into two types: *grouping-attribute* and *sorting-attribute*. The former indicates attributes that are used to group items, and the latter indicates the attributes that are used to sort items. We denote them as $\mathcal{P}_G \subseteq \mathcal{P}_F$ and $\mathcal{P}_S \subseteq \mathcal{P}_F$ respectively.

Here, let us denote intention-level designed relations as $L_{IL} = \{L_{IL}^{(1)}, \dots, L_{IL}^{(D)}\}$. As mentioned in Section 3.2.1, this study considers the following three types of designed relations: *parallel*, *contrast* and *ordinal*. The relations between items are determined using the sets of attributes, $\mathcal{P}_G, \mathcal{P}_S$, as follows.

parallel Two different items i and j have this relation when the items share one or more grouping-attributes; that is, $f_p(i) = f_p(j), \exists p \in \mathcal{P}_G$.

contrast Two different items i and j have this relation when the items do not share any grouping-attributes; that is, $f_p(i) \neq f_p(j), \forall p \in \mathcal{P}_G$.

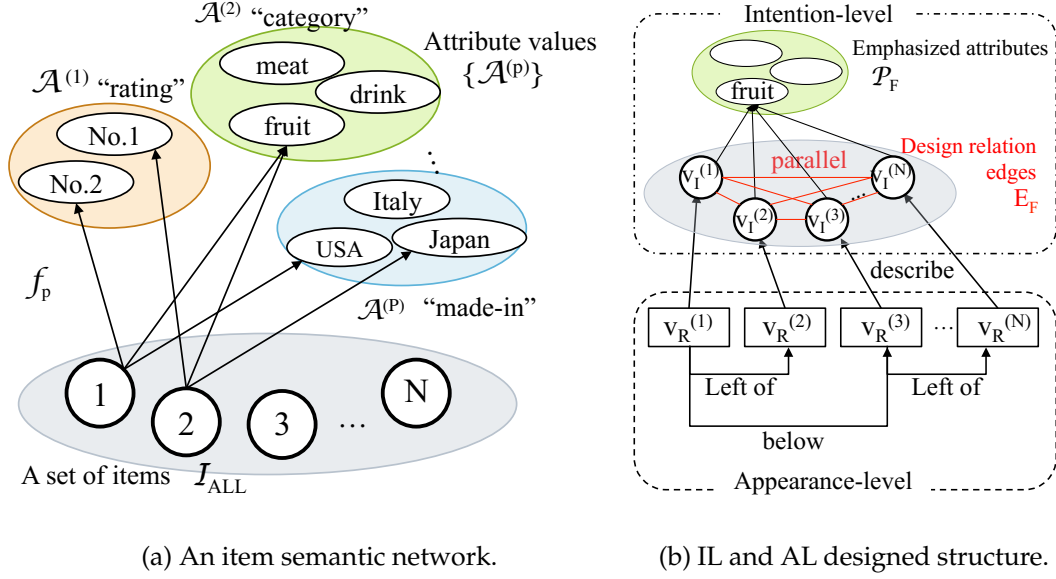


Figure 3.2: Graph-based representation of content structure.

ordinal Two different items i and j have this relation when the items share one or more grouping-attributes; that is, $f_p(i) = f_p(j), \exists p \in \mathcal{P}_G$, and are consecutive in an sorting-attribute $p' \in \mathcal{P}_S$.

We represent the relations among items as a directed graph. Let us denote a set of nodes $V_I = \{v_I^{(n)} | n = 1, \dots, N\}$ corresponding to the N items. Here, design-relation edges are defined as $E_{IL} \subseteq V_I \times V_I$, where each edge $e_d \in E_{IL}$ has one of design relations, L_{IL} . Finally, the intention-level designed structure is defined by a directed graph $G_{IL} = (V_I, E_{IL})$ (as shown in Figure 3.2 (b)).

Description of Appearance-level Designed Structure

Items in a digital catalog are described by various media that occupy spatial regions on the screen and the media regions are arranged based on a certain layout (appearance-level designed structure). In this study, we consider item regions composed of a set of media regions to be a basic region for the analysis. Let us denote a set of item regions constituting a visual catalog as $\mathcal{R} = \{R_1, \dots, R_N\}$ ($R_n \subset \Omega$, Ω : display space), where the information of n -th item is described by a region R_n . This study employs spatial relations among items to understand gaze transitions affected by the content layout. As the layout-oriented gaze transitions, we assume, for example, "looking at items from left to right" or "looking at items

along a row". Here, the layout of item regions is represented as several spatial relations $L_{AL} = \{L_{AL}^{(1)}, \dots, L_{AL}^{(D')}\}$ between every pair of item regions.

We also represent the appearance-level design relations among items as a directed graph for encoding eye movements. Let us denote a set of nodes $V_R = \{v_R^{(n)} | n = 1, \dots, N\}$ corresponding to the N item regions. Here, appearance design-relation edges are defined as $E_{AL} \subseteq V_R \times V_R$, where each edge $e_{AL} \in E_{AL}$ has one of design relations, L_{AL} . Finally, the appearance-level designed structure is defined by a directed graph $G_{AL} = (V_R, E_{AL})$.

3.4 Examples

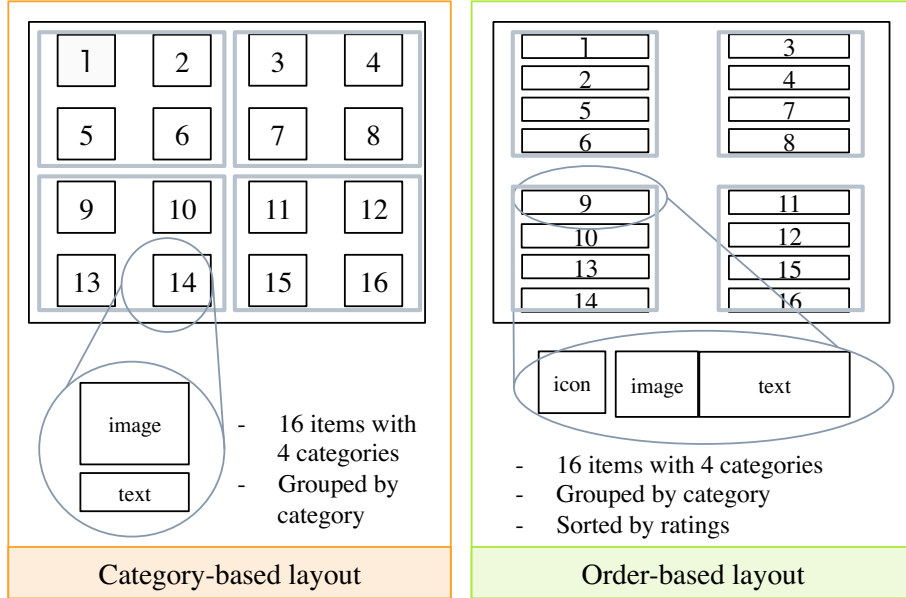


Figure 3.3: Examples of catalog layouts.

Here we give a few examples of designed structure of digital catalogs.

Example 1: category-based layout The layout is shown in Fig. 3.3 left. The catalog includes information about 16 different items ($N = 16$), and each item has two attributes $P_{All} = \{category, price\}$. Here, the set of emphasized attributes is $P_F = \{category\}$. The items are grouped according to their categories. The items in the same categories are linked by *parallel* relations and other items in different groups are linked by *contrast* relations. That is, design relations in this layout are $L_{IL} = \{parallel, contrast\}$. The intention-level designed structure of this catalog is shown in Fig. 3.4 (a).

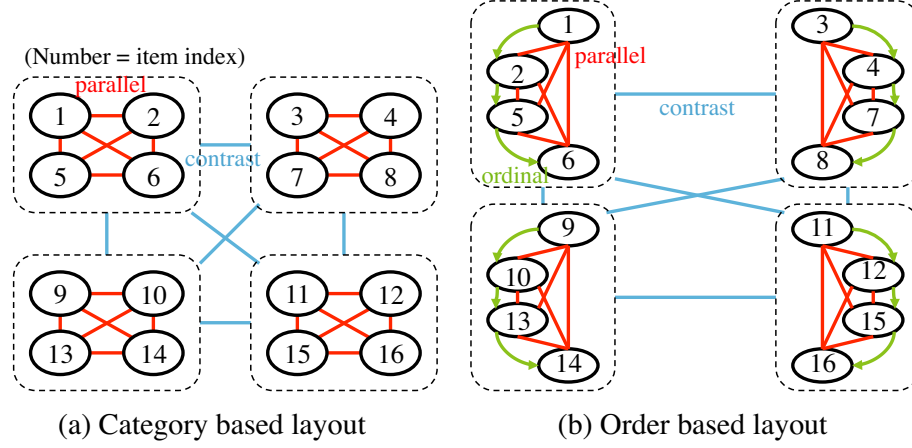


Figure 3.4: The intention-level designed structure of the example catalogs. Edges between dotted frames indicate that every node in the frame is connected to every node in the other frame.

Example 2: order based layout The layout is shown in Fig. 3.3 right. The catalog includes information about 16 different items ($N = 16$), and each item has two attributes $P_{All} = \{category, price, rating\}$. Here, the set of emphasized attributes is $P_F = \{category, rating\}$. The items are grouped according to their categories and sorted by their ratings in each group. Their ratings are described by an icon of a number from 1 to 4. Besides two design relations in the above example 1, items with successive ratings are linked with *ordinal* relations, i.e., $L_{IL} = \{parallel, contrast, ordinal\}$. The intention-level designed structure of this catalog is shown in Fig. 3.4 (b).

3.5 Discussion

In this section, we discuss the limitations of modeling designed structure of visual content for eye tracking analysis.

Extending the Representation of Designed Structure

In this chapter, we focus on simple designed structure with a few types of intention-level/appearance-level design relations among items. It is possible to extend the representation of designed structure to manage a greater variety of content structure, including complex hierarchical structure and dynamics (video content). For this, we would introduce a directed graph to represent various relations among content elements and consider temporal changes of designed struc-

ture. For appearance-level design structure, we would introduce more rich appearance information such as saliency and characteristics of the media.

Chapter 4

Encoding Eye Movements into Gaze Actions Using Designed Structure

4.1 Introduction

In this chapter, we propose a method to characterize eye movement patterns using designed structure defined in Chapter 3. The main subject of this chapter corresponds to the conversion from the second layer (unary gaze targets) to the third layer (multi-object gaze actions) in the hierarchical content browsing model in Chapter 2 (see Figure 2.2). That is, here we aim to define meaningful gaze actions that involve multiple objects on a screen based on sequences of unary gaze targets. Moreover, the gaze actions defined in this chapter will be employed to estimate content browsing states in the following Chapter 5 and Chapter 6 assuming frequent gaze actions may differ depending on which sub-task a browser is performing. In this chapter, the proposed gaze actions are evaluated by measuring classification accuracy of browsing states.

As described in Chapter 2, sequences of fixations/dwells (scan-paths) are commonly analyzed to understand underlying cognitive process of human browsers. However, it is difficult to interpret those sequential gaze data since their diversity is hard to manage. When we want to analyze content browsing behavior with a complex task, those sequences become longer and more diverse. To analyze sequential gaze data considering temporal information (i.e., the order of regions on a screen being looked at), some studies employ n-gram analysis to sequential eye tracking data. For example, dwell sequences annotated with semantic labels of areas-of-interest (AOIs) are split into tri-grams, and then frequent tri-gram patterns are discovered as meaningful gaze ac-

tions [Nakano and Ishii, 2010].

Although the n-gram analysis performs well in certain experimental settings, there are some limitations. First, gaze patterns can be very diverse if we directly use semantic labels of AOIs to represent gaze patterns. The definition of AOIs is normally task specific and heuristic due to the variety of semantic information of visual content. Therefore, appropriate labels and object region boundaries need to be defined by analysts for each content browsing situation or task. Second, human gaze is affected not only by what kind of objects the visual content contains but also by where and how they are displayed on the screen. Content appearance structure (e.g., spatial layouts of images and text) should be employed to characterize eye movements. The semantic and appearance structure in the visual content are highly related with each other. For instance, semantically similar objects are often placed close together in space. Therefore, we need to consider how to employ both of the structures jointly to understand content browsing behavior.

In this study, we present a framework to understand content browsing behavior by understanding which part of designed structure attracts browser focus and how behavior is influenced by the designed structure proposed in Chapter 3. Visual content such as web pages, digital catalogs and pamphlets usually have spatio-semantic structure designed by content creators. Content designers decide the positions and decorations of content elements (e.g., images and text) based on their perceptions and intentions (e.g., group items by their categories) and emphasize the relationships among the elements by design components (e.g., frames). Since content designers usually organize content structure to make the information in the content more comprehensible to browsers, we hypothesize that browsers' gaze behavior is highly affected by the designed structure. There are several benefits of characterizing eye movements with designed structure. First, it enables simple representation of semantic aspects of visual content; thus, it is expected to manage a variety of semantic and spatial structure of visual content. For example, the catalog structure can be described by only a few types of design relations (see Section 3.4). Second, designed structure is content independent. For example, as long as the designed structure is common, the same model/classifier can be used among different catalogs even their contents (a set of items) are different.

In the following sections, we describe the details of the proposed method to characterize gaze actions using designed structure of visual content. Then, we present an eye tracking experiment in a catalog browsing situation and its results

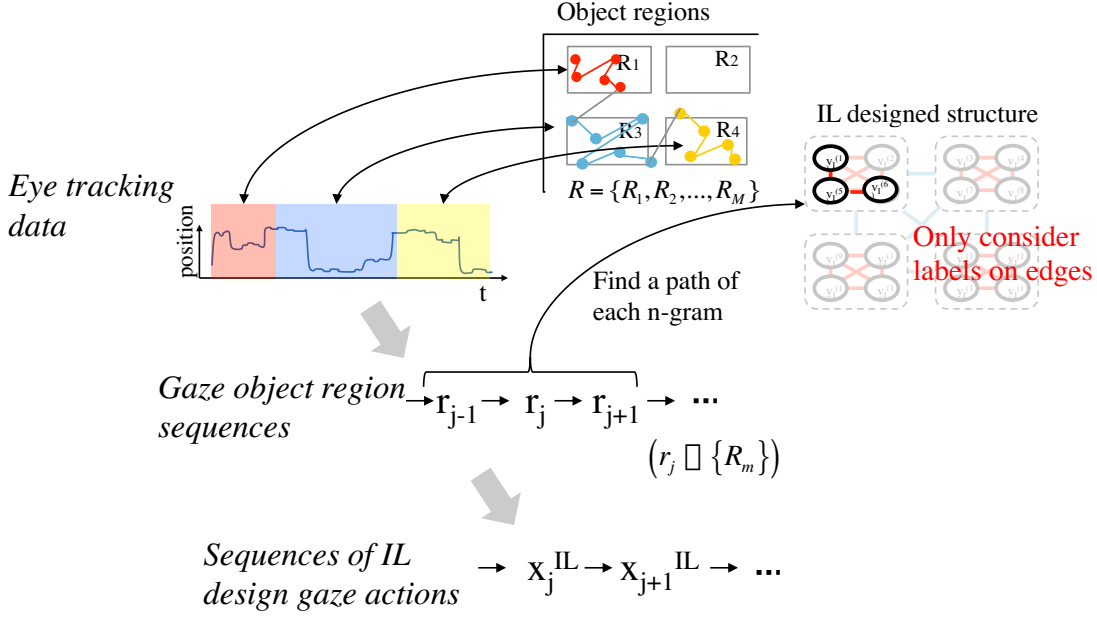


Figure 4.1: Labeling gaze transitions using designed structure.

to evaluate the effectiveness of the proposed method.

4.2 Proposed method

The overview of the proposed method of encoding eye movements into gaze actions is shown in Figure 4.1. Eye tracking data are first converted into sequences of object regions being looked at on a screen. Then, for each n-gram of gaze regions in the gaze region sequences, a path of corresponding item nodes in the graph representation of intention-level designed structure, G_{IL} , or appearance-level designed structure, G_{AL} , is obtained. The following sections assume the description of two levels of designed structure, G_{IL} and G_{AL} , are known for the analysis (see Chapter 3 for details).

In the following sections, we first describe the details of the proposed encoding method of eye movements into gaze actions in Section 4.2.1. Then, we present an analysis method of sequences of the encoded gaze actions in Section 4.2.2

4.2.1 Labeling Gaze Transitions Using Designed Structure

In a content browsing situation described in Section 1.3, browsers' eye tracking data are obtained as a sequence of gaze points on the screen, $E = (e_1, \dots, e_T)$ ($e_t \in$

Ω). Gaze targets (i.e., item regions being looked at) can be easily identified by associating each gaze point with an item region. Let us define a function $R : \Omega \rightarrow \mathcal{R}$, where $R(e) = R_n$ for $e \in R_n$. As the first step in interpreting gaze transition patterns, a sequence of timings of gaze transitions is obtained as $T = (t_1, \dots, t_J)$ by finding the timing $t_j \in \{2, \dots, T\}$ that satisfy $R(e_{t_{j-1}}) \neq R(e_{t_j})$. A sequence of regions being looked at is acquired as $r = (r_0, \dots, r_J)$ ($r_0 = R(e_1), r_j = R(e_{t_j})$). Here, the temporal interval U_j that r_j is being looked at is described as $U_j = [t_j, t_{j+1} - 1]$.

The gaze region sequence, r , is associated with design relation labels derived from the graphs of the designed structure, G_{IL} and G_{AL} . When the gaze actions derived from intention-level designed structure, they are called *IL design gaze actions*; meanwhile, ones derived from appearance-level designed structure are called *AL design gaze actions* in this study. Here, we describe bi-grams and tri-grams derived by intention-level designed structure as follows.

Bi-gram gaze actions The gaze transitions are associated with design relation labels derived from intention-level designed structure. For each gaze transition on timing t , the two item regions looked at, i, j , are found. According to the rules in Section. 3.3, the design relation between two items can be determined. Each gaze transition at t_j is represented by a D -dimensional vector $x_j^{IL} \in \{0, 1\}^D$, where the d -th element of x_j^{IL} is 1 if the two items has the d -th design relation label L_d . Finally, a sequence of annotated gaze transitions is obtained as $x^{IL} = (x_1^{IL}, \dots, x_J^{IL})$.

Tri-gram gaze actions For each tri-gram of gaze regions $(r_{j-1}, r_j, r_{j+1}) (j \in [1, J - 1])$ in the gaze region sequence, r , a path of corresponding item nodes in the graph G_{IL} is obtained. For simplicity, by considering only the labels of edges of each path, we can categorize the paths into six patterns as shown in Figure 4.2 (b). Note that the four top features in Figure 4.2 (b) correspond to tri-grams with three different gaze regions, that is, $r_{j-1} \neq r_{j+1}$, and the two bottom ones correspond to re-fixations, that is, $r_{j-1} = r_{j+1}$.

For AL design gaze actions, each gaze transition is labeled with their corresponding spatial relations among item regions. This study considers spatial directions from a given item region to its four neighbors, i.e., the set of spatial relations is as follows: $L_{AL} = \{\text{left-of}, \text{right-of}, \text{below}, \text{above}, \text{far}\}$. Gaze transitions between item regions that are not within four neighbour distance are labeled as *far*. The

appearance-level gaze action at t_j is represented by one-of-K representation, i.e., the feature is denoted as a $|L_{AL}|$ -dimensional vector $x_{jk}^{AL} \in \{0, 1\}^{|L_{AL}|}$, where one of the elements corresponding to the spatial relation at the timing t_j is 1 and all other elements are 0.

Eye trackers sometimes contain noise or miss browsers' gaze points because of blinks. Therefore, the sequence of gaze points is first smoothed by applying a median filter¹. Moreover, the sequence of regions looked at is modified by discarding intervals shorter than a threshold. In the following experiment, 100 ms is used as the threshold. If successive intervals with the same item ID are interrupted by a blank, the intervals are combined to a longer interval.

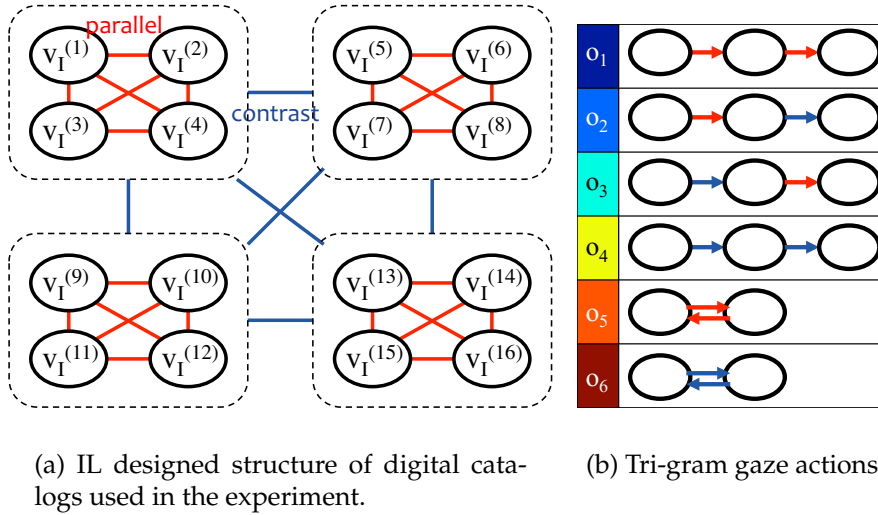


Figure 4.2: Encoding of eye movements. (a) Edges between dotted frames indicate that every node in the frame is connected to every node in the other frame. (b) Six patterns of tri-gram paths in the graph G_{IL} . The two bottom features correspond to comparison patterns in which the first and the last dwell of the tri-gram is on the same item.

4.2.2 Analysis of Labeled Gaze Transitions

For the later analysis in this chapter, we only use bi-gram patterns for simplicity. For estimating browsing states in Chapter 5 and 6, we use tri-gram patterns since it can represent re-fixation patterns, which may be important to understand browsing states in catalog browsing situations.

¹In this thesis, the window size of the median filter is 5 sampling points at 60 Hz (corresponding to about 83 msec).

As a result of the interpretation, browsers' gaze behavior is represented as multiple sequences of vectors, $\mathbf{x}^{IL}$ and $\mathbf{x}^{AL}$. The frequency distributions of features are calculated with each interpreted gaze transition sequence ($\mathbf{x}^{IL}$ and $\mathbf{x}^{AL}$) as

$$\mathbf{X}^{IL} = \sum_j \mathbf{x}_j^{IL} / J \quad (4.1)$$

$$\mathbf{X}^{AL} = \sum_j \mathbf{x}_j^{AL} / J. \quad (4.2)$$

4.3 Experiment

In this experiment, we evaluate the proposed framework for encoding eye movements into gaze actions using designed structure of visual content. Assuming different browsing states induce different types of gaze actions, the purpose of the experiment is to assess which encoded gaze actions is more effective to represent differences among browsing states when combined with motion features of eye movements. Since the ground truth data of browsing states are not available, we instead gave participants a set of objectives that correspond to browsing states in choice behavior. Then, the proposed framework is measured by the performance of estimating browsers' objective from their gaze. This section presents the experimental methodology and the results.

4.3.1 Experimental Settings

12 participants took part in the experiment. Each participant was asked to sit in front of a screen showing a digital catalog (see Figure 4.3). Eye tracking data of the participants were acquired as 2-d points on the screen by using an eye tracker² installed below the screen.

Digital catalogs. To investigate the general versatility of the proposed framework, we used digital catalogs with two different types of layouts: category-based and order-based (see Figure 3.3). The details of each layout are described in Section 3.4. With each layout, four digital catalogs are prepared. Each digital catalog contained the description (images and text) of 16 items, and the items can

²Tobii X120 (freedom of head movement: 400x220x300 mm, sampling rate: 60 Hz, accuracy: 0.5 degrees)



Figure 4.3: The experimental environment.

be grouped into one of four categories: accessories, home electronics, house-hold goods, and toys.

Objectives of browsers. The participants were asked to select a gift for friends/acquaintances. According to existing studies in the marketing research field, the choice process can be divided into a few stages such as *orientation* and *evaluation* [Payne, 1976, Russo and Leclerc, 1994, Gidlöf et al., 2013]. In this experiment, we assume three stages and use three objectives which correspond to “orientation stages”, where a browser is obtaining the overview of the catalog, “evaluation stages”, where a browser is actually comparing candidates to make a selection, and “post-selection stages”, where a browser is free-viewing or confirming their decision, respectively. Specifically, the following series of objectives were given to each participant, which we call *input*, *decision* and *free-viewing*.

Input: (30 sec) *Browse a digital catalog to confirm what products exist.*

Decision: (no limit) *Select a gift from the catalog for a designated person considering his/her profile.*

Free-viewing: (60 sec) *Browse the catalog freely.*

If the participants were to select a gift for their real friends/acquaintances, it is possible that the personality of the recipient of the gift or the relationship between the participant and the recipient can affect the participant’s behavior. Since such information is hard to be acquired through the experiment, we designate the recipient by showing the profile of a certain person. The provided profile includes the information of the fictitious relationship with the browser (i.e., the

participant), the hobby, and the portrait. The profile was displayed to participants between *input* and *decision* objectives.

4.3.2 Comparative Eye Movement Features

Unary Gaze Targets

To evaluate the effectiveness of using designed structure to understand browsing behavior, we prepare features that use the item-semantic attributes of gaze targets as comparative features. Let us denote a set of attribute values of the gaze target item at the timing t_j as $\mathcal{S}^{(n)} = \{f_1(n), \dots, f_p(n)\}$ ($n \in \{1, \dots, N\}$). Here, the attribute values, $\mathcal{S}^{(n)}$, are also described as one-of-K representation, i.e., the p -th attribute is denoted as a Q_p -dimensional vector $\mathbf{x}_{jp}^S \in \{0, 1\}^{Q_p}$, where one of the elements corresponding to $f_p^{(n)}$ is 1 and all other elements are 0. The vector representing all semantic attributes is obtained by combining the vectors of attributes as $\mathbf{x}_j^S = (\mathbf{x}_{j1}^{ST}, \dots, \mathbf{x}_{jP}^{ST})^T$. Finally, we obtain a sequence of semantic attributes looked at as $\mathbf{x}^S = (\mathbf{x}_1^S, \dots, \mathbf{x}_J^S)$.

Motion Features of Eye Movements

This study uses motion features of eye movements from [Sugano et al., 2014] including fixation features (positions, duration, and time) and saccade features (directions, length, duration, and time). Although the motion features of eye movements for each right and left half of the screen are defined separately in [Sugano et al., 2014], we do not distinguish them in this thesis. Moreover, note that the features related to the time information are removed in the following objective classification. This is because the objective classification would be obvious if the time information is used since the objectives in the experiment are sequentially given to the participants. In addition to the motion features of eye movements, the durations of regions being looked at are also examined. As described in the previous section, the duration of each gaze region r_j is denoted as $|U_j| = t_{j+1} - t_j$. We use the mean value and variance of the durations $\{|U_j|\}$. The motion features of eye movements are described as a vector $\mathbf{X}^{Gaz}$. All the features used in this thesis are listed in Table 4.1.

4.3.3 Results

Objective Classification

As explained above, the proposed framework is evaluated by measuring the performance of classification of the given objectives. We classify participants' gaze sequences based on the frequency distributions of eye movement features calculated in Section 4.2.2 and Section 4.3.2 in a supervised learning manner. In the experiment, we investigate which content feature is more effective when combined with motion features of eye movements. For motion features of eye movements, in addition to fixation features and saccade features from [Sugano et al., 2014], the mean value and variance of duration lengths between gaze transitions are used. For the details on the motion features of eye movements in the experiment, see Section 4.3.2 and Table 4.1.

For the classification algorithm, we used Random Forest [Breiman, 2001]. Random Forest is an ensemble learning method using a set of decision trees. At the training phase, the Random Forest algorithm builds each decision tree using a subset of training data via random sampling with replacement. After training, an unseen feature vector can be classified based on a majority vote from the learned trees. Random Forest is known as a robust and fast learning method and it allows us to assess the importance of feature variables. After the training phase, a part of the training data would be left without being used. Using the out-of-bag (OOB) data as test data, an unbiased classification accuracy can be obtained. The OOB importance of the n -th variable can be calculated by measuring the difference between the original classification accuracy and the accuracy obtained by randomly permuting the n -th variable values in the OOB data.

Finally, we obtained 144 gaze sequences (12 participants $\times$ 4 catalogs $\times$ 3 objectives) with each layout type of catalog. The features and their ID numbers are listed in Table 4.1. For item-semantic features, we consider the categories and price ranges of items for the category-based layout. For order-based layout, we consider the ratings of items in addition to the price ranges and categories.

Classification accuracies (precisions) were obtained by leave-one-subject-out cross-validation, i.e., gaze data of one subject was used as test data and the rest of the gaze data were used as training. The results are shown in Table 4.2. With both category-based and order-based layouts, the highest classification accuracy is obtained with the combination of motion features of eye movements and proposed gaze actions based on both intention-level and appearance-level design structure.

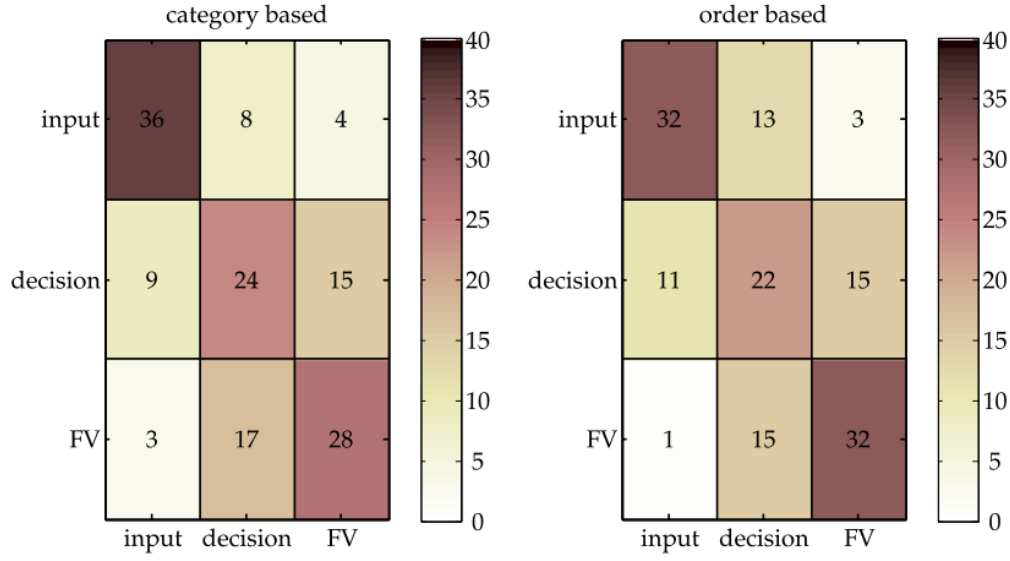


Figure 4.4: The confusion matrices of the classification of the objectives. The horizontal axis indicates classified objectives and the vertical axis indicates actual objectives. Each number in the matrices indicates the number of gaze sequences.

The results show that using content design information to interpret browsing behavior is more effective than compared with item-semantic features. The use of both IL and AL design gaze actions improves the classification accuracy compared to when only one of them is used, which indicates that both designers' intentions and content appearance are important factors in the browser objective classification.

Moreover, we examined the confusion matrices of the three-class classification to investigate the separability between the objectives (Figure 4.4). With both layouts, the *input* objective is classified with the highest accuracy. It can also be seen by comparing the F measures in Table 4.2. For the *input* task, it is possible that participants do not have a clear intention for content-browsing because it occurs before they see a profile of the target of selecting a gift; meanwhile, for the *decision* task and the *free-viewing* task, participants browse the digital catalogs based on their own interests/intentions. The results indicate that browsing states are more separable in the proposed method if there is a huge gap in degrees of browsers' sense of purpose and/or intentions.

The OOB feature importance is obtained using the 144 sequences of gaze data combining motion features of eye movements with all of the content features (item-semantic features, AL design gaze actions and IL design gaze actions). The

OOB feature importance with category-based layouts is shown in Figure 4.5, and one with order-based layouts is shown in Figure 4.6. In Figure 4.5, with category based layouts, the results show that intention-level gaze actions (feature ID: 14 and 15) has the most importance among other features. Among motion features of eye movements, mean and variance of the duration of each gaze region (feature ID: 16 and 17, respectively), variance of fixation positions (feature ID: 20 and 21) and variance of fixation duration (feature ID: 24) contribute to the estimation. In Figure 4.6, with the order based layout, one of variables of intention-level gaze actions: the frequency of gaze transitions that follow *ordinal* relations (feature ID: 20), has the highest importance as well. Moreover, it also shows that horizontal spatial relations in appearance-level gaze actions (feature ID: 16 and 17) contribute to the estimation. These results indicate the effectiveness of the proposed gaze actions to represent content-browsing behavior with both content layouts compared to other features.

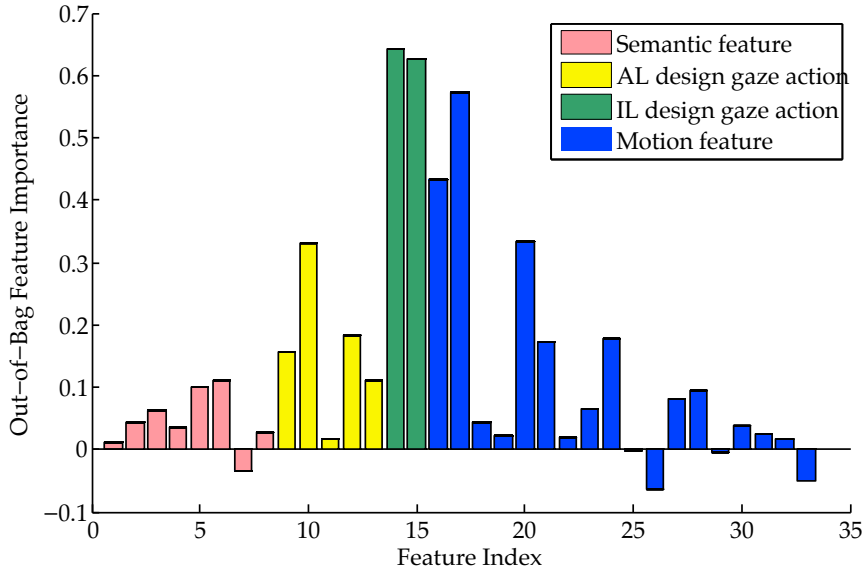


Figure 4.5: Out of bag feature importance with category based layout. The feature IDs of horizontal axis correspond to feature ID listed in Table 4.1. The color of bars indicates the type of features.

4.4 Discussion

In summary, this chapter proposed a method to encode eye movements into gaze actions using designed structure of visual content. We confirmed the effectiveness

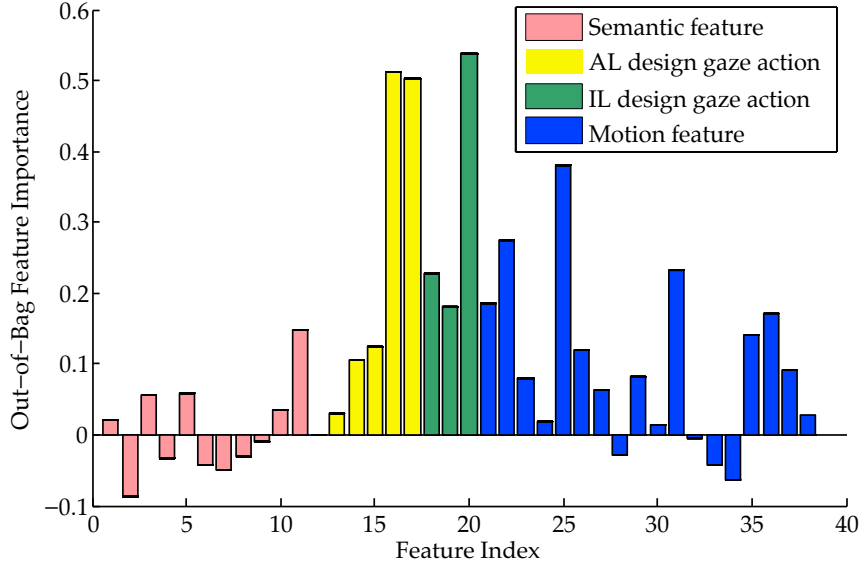


Figure 4.6: Out of bag feature importance with order based layout. The feature IDs of horizontal axis correspond to feature ID listed in Table 4.1. The color of bars indicates the type of features.

of the gaze actions to represent differences among browsing states through the experiment through the eye tracking experiment. The results showed that the use of the gaze actions based on both IL and AL designed structure improved the classification accuracy of participants' objectives.

In the following sections, we discuss the limitations and future work of the proposed framework.

The Limitation of Our Approach

Although the experimental results showed that designed structure is useful to interpret browsing behavior, there is a limitation that it can only deal with the situation where the browser understands designed structure by looking at visual content. That is, if the appearance of the visual content does not reflect its designed structure in a comprehensible way to browsers, designed structure has less impact on browsers' gaze behavior. For example, in the category based layout, items with *parallel* relations are surrounded with a frame (see Figure 3.3). As a preliminary experiment, we investigated the effects of the representation way of designed structure using category based layouts without frames. Using the catalog without frames, the accuracy of task estimation was 0.527 using the combination of the motion features of eye movements and both of the gaze actions

(i.e., design oriented features), which is lower compared to the results shown in Table 4.2. The comparison implies that the proposed method can deal with only gaze behavior with well-designed visual content.

Table 4.1: Features used in the experiment. The left column shows the name of features. The second and third column from the left shows feature ID with category-based layout (CL ID) and feature ID with order-based layout (OL ID), respectively. The right column shows the particular labels/reasons/values used in the experiment.

	CL ID	OL ID	Features
Item-semantic Feature (price range)	1	1	20,000- yen
	2	2	10,000-15,000 yen
	3	3	5,000-7,000 yen
	4	4	1,000-3000 yen
(Category)	5	5	Accessories
	6	6	Home electronics
	7	7	House-hold goods
	8	8	Toys
(Rating)	N/A	9	First
	N/A	10	Second
	N/A	11	Third
	N/A	12	Fourth
AL Design Gaze Action (positions)	9	13	Far
	10	14	Left-of
	11	15	Right-of
	12	16	Below
	13	17	Above
IL Design Gaze Action	14	18	Contrast
	15	19	Parallel
	N/A	20	Order
Motion Feature (duration)	16	21	Mean
	17	22	Variance
(Fixation Position)	18	23	Mean (Horizontal)
	19	24	Mean (Vertical)
	20	25	Variance (Horizontal)
	21	26	Variance (Vertical)
	22	27	Covariance
(Fixation Duration)	23	28	Mean
	24	29	Variance
(Saccade Direction)	25	30	Mean (Horizontal)
	26	31	Mean (Vertical)
	27	32	Variance (Horizontal)
	28	33	Variance (Vertical)
	29	34	Covariance
(Saccade Length)	30	35	Mean
	31	36	Variance
(Saccade Duration)	32	37	Mean
	33	38	Variance

Table 4.2: The estimation accuracies (precisions) and F measures of the three objectives: input (IN), decision (DE) and free-viewing (FV). The 6 rows from the top show the results with the category-based layout, and the next 6 rows show the results with the order based layout. The first row shows the results obtained using only motion features of eye movements (Gaze). The following four rows show the accuracy obtained combining motion features of eye movements with item-semantic features (Sem), appearance-level (AL) design gaze actions, intention-level (IL) design gaze actions and two content-design-oriented features (AL and IL design gaze actions), respectively. The 6th row shows the accuracy obtained combining motion features of eye movements with all content features: item-semantic features, AL design gaze actions and IL design gaze actions.

	Accuracies				F measure		
	Avg.	IN	DE	FV	IN	DE	FV
Layout Type: Category based							
Gaze	0.583	0.667	0.542	0.542	0.674	0.500	0.584
Gaze+Sem	0.535	0.563	0.521	0.521	0.607	0.459	0.556
Gaze+AL	0.597	0.750	0.521	0.521	0.758	0.485	0.556
Gaze+IL	0.604	0.709	0.563	0.542	0.731	0.529	0.559
Gaze+AL+IL	0.611	0.750	0.500	0.583	0.750	0.495	0.589
Gaze+All	0.604	0.729	0.521	0.563	0.737	0.495	0.587
Layout Type: Order based							
Gaze	0.514	0.583	0.458	0.500	0.615	0.411	0.533
Gaze+Sem	0.556	0.625	0.438	0.604	0.652	0.433	0.586
Gaze+AL	0.576	0.729	0.417	0.583	0.729	0.417	0.583
Gaze+IL	0.583	0.688	0.396	0.667	0.673	0.409	0.660
Gaze+AL + IL	0.597	0.667	0.458	0.667	0.696	0.449	0.653
Gaze+All	0.569	0.667	0.375	0.667	0.667	0.391	0.640

Chapter 5

Unsupervised Estimation of Content Browsing States

5.1 Introduction

In this chapter, we present a probabilistic model to represent transitions of content browsing states. The main subject of this chapter corresponds to the conversion from the third layer (multi-object gaze actions) to the fourth layer (browsing states) in the hierarchical content browsing model in Chapter 2 (see Figure 2.2). Here we aim to estimate browsing states based on gaze actions while assuming browsing states can be characterized by the gaze actions defined in Chapter 4. Unlike Chapter 4, where we assumed that the browsing states of human browsers are known and can be categorized into a few pre-defined states, this chapter proposes an estimation method of unknown browsing states in a bottom up manner.

As described in Chapter 2, we assume that a goal-directed browser split their task/goal into a set of sub-tasks that are easier to accomplish. Each browsing state in this study corresponds to one of the sub-tasks, which may change over time. When the number and types of browsing states are unknown and not observable, a state space model is often used to estimate the hidden states from observable valuables. This study takes this approach and employ a state space model to represent transitions of browsing states. Simola et al. [Simola et al., 2008] take a similar approach to ours: they apply a hidden Markov model (HMM) to sequences of motion features of eye movements (e.g., saccade lengths) to uncover processing states of information-search in sentence reading. Three HMMs were built corresponding to three different information-search tasks: simple word search, finding the answer to a question, and selecting an interesting title from a list. Their results

showed that the tasks can be estimated from a newly observed sequence of eye movements by comparing the likelihood of each model.

However, there are some problems with using HMMs for modeling other content browsing situations where the amount of time a viewer spends in a state is an important feature. Therefore, we employ a hidden semi-Markov model (HSMM) so that we can estimate the browsing states from eye tracking data in a bottom-up manner with their duration distribution. The HSMM is an extended model of the HMM that each state has a duration distribution that is explicitly specified. HSMM is often used to model human behavior in various contexts.

To evaluate how the proposed model estimates meaningful hidden browsing states behind the eye movement sequences, we collected eye tracking data in a catalog browsing situation with a procedure explained in Section 5.3. The data is later used for the training of the model (Section 5.2.3) and its evaluation (Section 5.3.3).

5.2 Modeling and Estimating of Content Browsing States Using Hidden Semi-Markov Model

The overview of the proposed model is shown in Figure 5.1. The observed eye movements are first encoded into gaze actions using designed structure of visual content (see Chapter 4). Then a HSMM is applied to the sequence of encoded gaze actions to model probabilistic relations among gaze actions and browsing states (Section 5.2.3).

5.2.1 Sequence of Gaze Actions

This chapter assumes that the observed eye tracking data can be converted to a sequence of gaze actions using the method presented in Chapter 4. Two levels of designed structure were defined in Chapter 3: intention-level (IL) and appearance-level (AL) designed structure. The experimental results in Chapter 4 have shown that gaze actions using IL designed structure were the most effective to represent differences among given objectives (browsing states) of browsers. Thus, in this chapter, we focus on gaze actions based on IL designed structure.

In Chapter 4 the two types of encoding eye movement were presented: bi-gram gaze actions and tri-gram gaze actions (see Section 4.2.1). Although we consider only bi-grams in Chapter 4 when comparing with other features, tri-

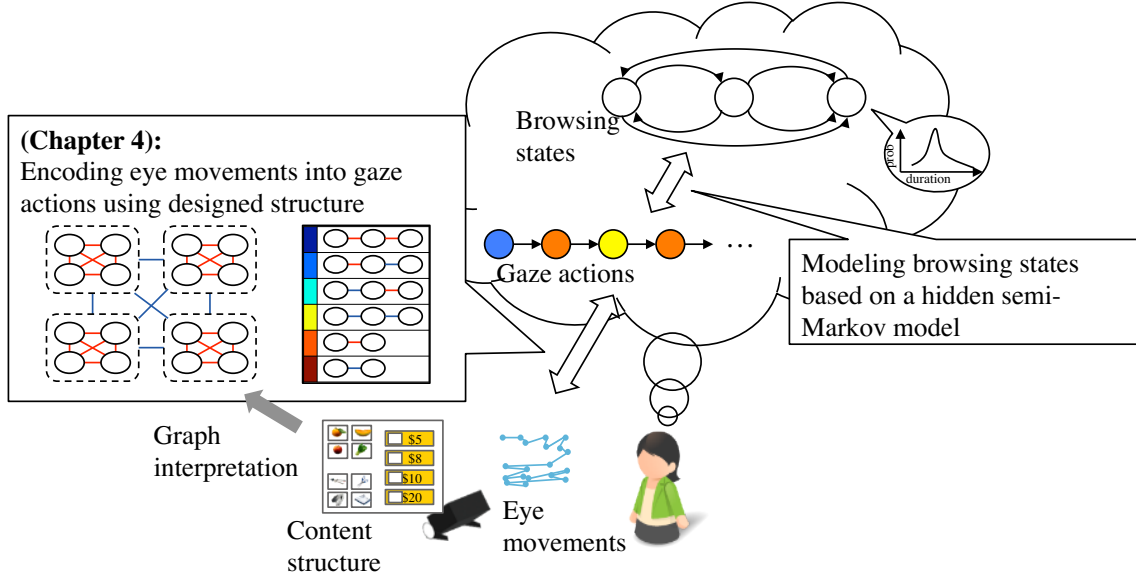


Figure 5.1: Overview of the proposed method.

grams are expected to be more informative to represent content browsing behavior. Thus, we employ tri-gram gaze actions for the rest of the analysis in this chapter. Note that the possible tri-gram gaze actions based on intentional-level designed structure are of six types, which are listed in Figure 4.2 (b).

5.2.2 Hidden Semi-Markov Model

A HSMM (also known as explicit duration hidden Markov model) is an extension of a HMM that considers an additional parameter for explicitly modeling the duration of states. Let us denote the discrete-output HMM as $\lambda_{\text{HMM}} = (\mathcal{S}, \mathcal{O}, A, B, \pi)$. $\mathcal{O}$ is a set of possible observations: $\mathcal{O} = \{o_1, \dots, o_6\}$, which correspond to the set of tri-gram gaze actions in Figure 4.2 (b). $\mathcal{S}$ is a set of hidden states: $\mathcal{S} = \{s_1, \dots, s_M\}$. A is a $M \times M$ state transition probability matrix, B is a $M \times 6$ emission probability matrix, and π is a set of initial probabilities of the states.

The HSMM used in this thesis can be represented as $\lambda_{\text{HSMM}} = (\mathcal{S}, \mathcal{O}, A, B, P, \pi)$ by adding a $M \times D$ duration distribution matrix P to the original HMM, where D is the maximum state duration and $p_{m,d} \in [0, 1]$ indicates the probability that the m -th state last for d -time units.

5.2.3 Learning Model Parameters

For learning the model parameters, this study uses the efficient EM algorithm proposed in [Yu and Kobayashi, 2006]. Since the available data is limited, the duration distribution of states is approximated by a Gaussian distribution. Note that the duration distribution is a discrete distribution which is defined between 1 to D . The duration distribution of each state is normalized so that the sum of each row of the duration matrix, P , is 1.

In the EM iteration, uniform distributions are used for the initial values of A and π in the HMM and A , P , and π in the HSMM. For the initial values of B in both the HMM and the HSMM, distributions of the gaze actions are first obtained by applying a sliding window of size 20. Then, each mean point of the M distribution clusters obtained by k -means clustering are used.

To decide the number of states M for the HSMM and the HMM, we utilized the Akaike Information Criterion (AIC) since the sample size is limited against the model complexity [Cartella et al., 2015]. AIC can be represented as $AIC = -2 \ln L(\hat{\lambda}) + 2h$, where $\log L(\hat{\lambda})$ is the log likelihood of the model with the estimated parameters $\hat{\lambda}$, and h is the degree of freedom of the model parameters. The best number of states can be decided by finding M that minimizes the value of AIC.

5.3 Experiment

Eight participants took part in the experiment. Each participant was asked to sit in front of a display showing a digital catalog (see Figure 5.2). Gaze data of the participants were acquired as 2D points on the display by using an eye tracker¹ installed below the display.

Unlike the experiment in Chapter 4, here we aim to model browsing states including their temporal changes while performing a task. Accordingly, we did not give objectives to participants that correspond to browsing states. Eye movements were recorded in each session as a single sequence.

¹Tobii X120 (freedom of head movement: 400x220x300 mm, sampling rate: 60 Hz, accuracy: 0.5 degrees)

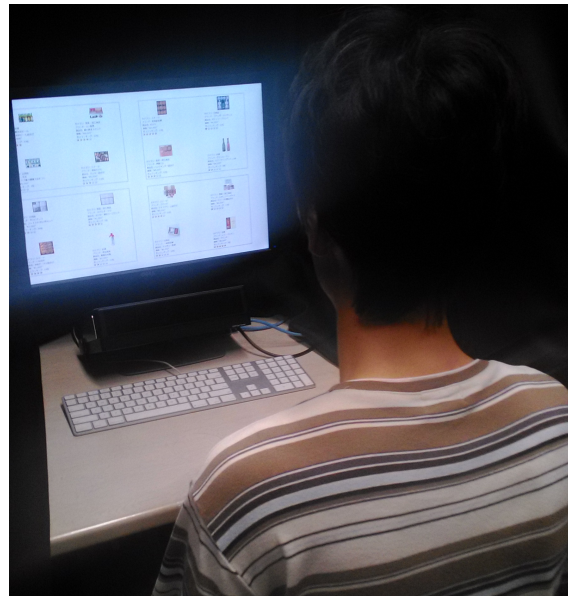


Figure 5.2: Experimental environment. A participant is browsing a digital catalog on a display. Gaze data of the participant are recorded by the eye tracker installed below the display.

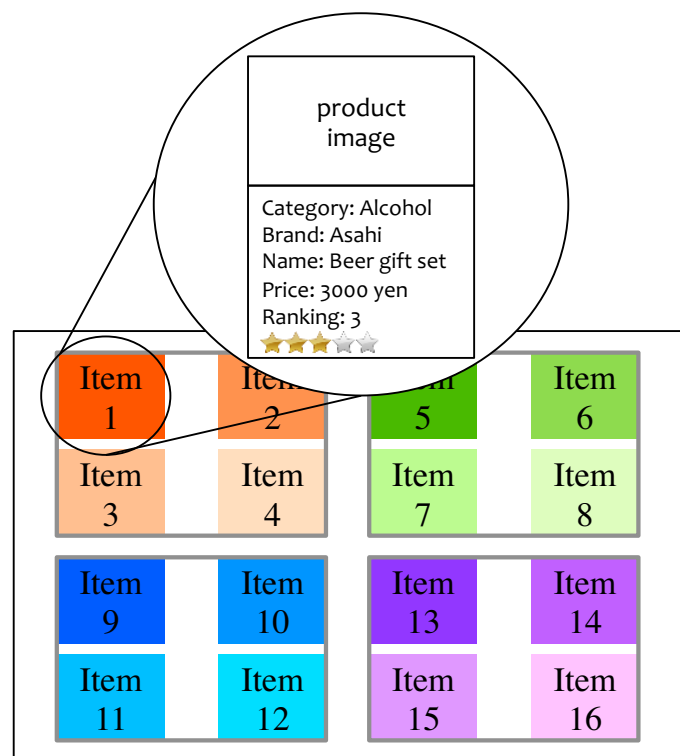


Figure 5.3: Layout of digital catalogs. Items with a same color hue are in a same group, and each group of items are divided by a gray frame.

5.3.1 Experimental Setting

Digital Catalogs

For each participant, eight digital catalogs were prepared. Each digital catalog contained the description (images and text) of 16 items (see Figure 5.3). The semantic attributes and attribute values of items that were available in this study are listed in Table 5.1. The semantic information was described in text on the catalogs so that browsers could understand it without prior knowledge. The items in each catalog were grouped by either *price* or *category* attribute. The item positions within a group were randomized every time the catalog was shown to a participant.

Table 5.1: The attributes and attribute values used in the experiments.

Category	Price (yen)	Ranking	Review
Delicatessen	1001-3000	1-4th	1-star
Sweets	3001-5000	11-14th	2-star
Alcohol	5001-7000	21-24th	3-star
Household goods	7001-	31-34th	4-star 5-star

Task Complexity

Participants were requested to select an item from a catalog as a seasonal gift. The complexities of decision tasks can affect browsers behavior [Payne, 1976]. We hypothesize that the task complexities may influence transitions and duration of browsing states as well. Such task complexity would vary based on a browser's preference and which products were available in the catalog. For example, it is possible that browsers would end up making a decision among multiple strong candidates, but it is also possible that they could only find one product that satisfies their criteria and thus the evaluation stage would be skipped. Pfeiffer et al. aimed to control the decision tasks by fixing the requirements of items to chose so that only one item satisfies all of the conditions to induce goal-directed search behavior [Pfeiffer et al., 2014].

We take a similar approach here; we gave participants requests that specified the requirements for items to select for each trial, for example: "Select an item in the *Alcohol* category and with more than a 4-star review." Differing from Pfeiffer's work, we change the number of items that satisfy the requests to be between 0 to

3 to investigate the relationship between browsing states and task complexities. In the rest of this thesis, the term *candidates* is used to indicate items that satisfy the experimenter's request. The number of gaze sequences recorded in the experiments with each condition of task complexity (number of candidates) is shown in Table 5.2. We discarded gaze sequences in which a participant chose an item that did not satisfy the given request (even though at least one candidate existed in the catalog). Our hypothesis about browsing states and task complexities is evaluated in Chapter 6.

Procedure

The first decision making session of each participant started with a short, 5-point calibration phase. Each participant performed eight decision making sessions. In each session, one of the requests was shown to the participant before showing the catalog. The participants were instructed to follow the request to the best of their ability. During catalog browsing, the participants were able to refer to the request by pressing the space key on a provided keyboard in case where they forgot the request. The gaze data during catalog browsing were registered together with the time stamps when the space key was pressed. The participants were instructed to press the enter key on the keyboard when they made a decision. When the enter key was pressed, the catalog on the monitor was suddenly hidden. After the catalog was hidden, the participant was asked to name the selected item.

Table 5.2: The number of recorded gaze sequences with each task complexity.

Number of candidates	0	1	2	3	Total
Number of sequences	16	15	15	16	62

5.3.2 Examples of Collected Gaze Data

Eye trackers sometimes contain noise, which can complicate analysis. To manage error, our sequences of gaze points were first smoothed by applying a median filter ². For the subsequent analysis, each sequence of gaze points was first converted to a sequence of dwells on items. Recall that a dwell is a set of successive fixations in each area of interest (AOI) and dwell sequence analysis was often used in previous studies on consumer decision analysis

²In this paper, the window size of the median filter was 5 sampling points at 60 Hz (corresponding to about 83 msec).

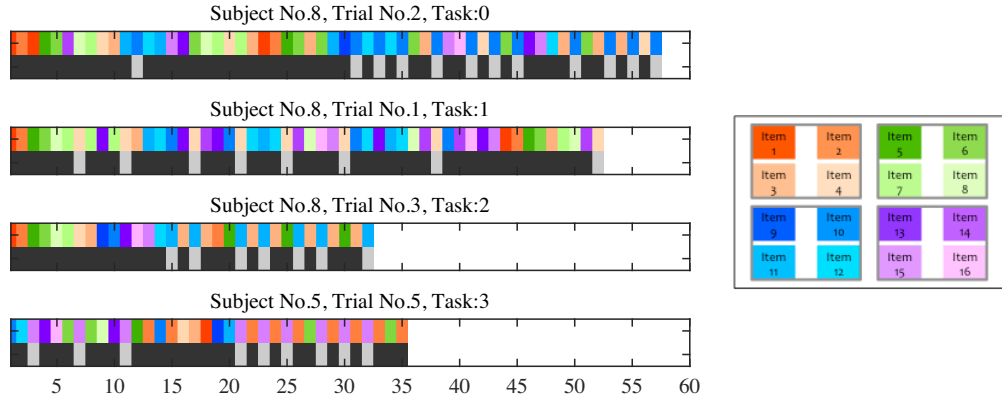


Figure 5.4: Example of dwell sequences. For each sequence, the color in the top row corresponds to the color of regions in Figure 5.3, and the bottom row shows the timings of dwells on the selected item by light gray highlights.

[Glaholt and Reingold, 2011]. In this study, AOIs are the regions of items and each dwell is simply identified by a set of successive sampling points in each item region without identifying fixations. The dwell sequences are modified by discarding dwells with duration shorter than a threshold (100 msec). If successive dwells with the same item ID were interrupted by a blink, the intervals were concatenated to a longer dwell. Any dwell in which the participant was examining the requirements of the request was also discarded.

Examples of dwell sequences are shown in Figure 5.4. In the examples, we can observe temporal differences in dwell frequencies on a selected item. In the top and the third examples, similar behavior to the gaze cascade effect [Shimojo et al., 2003] can be observed. That is, the selected items are more frequently looked at as time progresses. This second example supports the assertion by Russo et al. [Russo and Leclerc, 1994].

We can see an example of exploratory browsing behavior after the 138th dwell on the selected item, which can be interpreted as the verification stage. The bottom example highlights the importance of considering probabilistic transitions between decision stages. In the bottom example, there is an interval that contains fewer dwells on the selected item from the 12th dwell to the 20th dwell. Considering whether or not each item has been looked at, this example may be interpreted as follows: the participant was exploring the catalog until the 20th dwell with the interruption of evaluating behavior occurring at the beginning.

5.3.3 Results

Ideally, ground truth of the browsers' states could be used to evaluate the validity of the proposed model. Since ground-truth data were not available, we instead compared the estimated browsing states with findings from other choice behavior studies. Specifically, we evaluated the proposed model by its performance of identifying *evaluation* states [Russo and Leclerc, 1994], where strong candidates are compared and evaluated. Moreover, we attempted to use the estimated states to improve browsers' interest estimation. In the following sections, we first discuss the interpretation of the estimated browsing states using the collected gaze data in Section 5.3. Then, we show how the states can be used for browser's interest estimation.

Interpretation of Estimated Browsing States

We trained HSMMs with different numbers of hidden states using all the collected gaze data. The values of AIC are shown in Figure 5.5. Accordingly, the number of states in the HSMM, M , was selected; $M = 4$ (see Section 5.2.3 for details). The topology of the trained HSMM, A and π , and the output probabilities, B , are shown in Figure 5.6 and Figure 5.7, respectively. Figure 5.6 suggests that a browser tends to start in state s_1 or s_3 . Since transition probabilities to state s_3 or state s_2 are comparatively higher than ones to other states, a browser is likely to shift to the two states. From Figure 5.7, state s_1 or s_3 can be interpreted as a state in which a browser is more likely to look at items in the same group. The duration probabilities of the trained HSMM, P , is shown in Figure 5.8. The results show that the state s_2 is comparatively shorter than other states. To see the validity of the approximation by a Gaussian distribution, duration distributions of each state before fitting a Gaussian distribution are calculated as in Figure 5.9. The results suggest that we might need to employ a more complex distribution to approximate the duration distribution. For example, state s_1 shows the tendency of diphasic distribution. However, we need more data to decide which types of distribution should be used.

We investigated how the estimated browsing states can identify evaluation states where browsers compare strong candidates and the selected item. Figure 5.10 and Figure 5.11 show the ratio of dwells on the selected items, non-selected candidates, and others at in each estimated browsing state when the number of candidates was 2 or 3, respectively. In the both figures, browsers

are more likely to look at the selected items or candidates in the states s_2 and s_4 compared to in the states s_1 and s_3 . To associate the obtained states with the three-stage model [Russo and Leclerc, 1994], we consider states s_1 and s_3 to be *orientation* or *verification* (these two stages share a similar pattern, but can be discriminated based on context), while states s_2 and s_4 can be interpreted as *evaluation*. Moreover, from the estimated transition probabilities (Figure 5.6), a browser is more likely in states s_3 and s_2 . Thus, the two states, states s_3 and s_2 , can be interpreted as a representative state of *orientation/verification* stage and of *evaluation* stage, respectively.

An example sequence of posterior probabilities of browsing states is shown in Figure 5.12. To estimate the sequence of browsing states at each dwell, leave-one-subject-out cross validation was used; that is, gaze data of one participant was used as test data and the remaining data were used to train the model. The number of states of HMM, M_{HMM} , were also estimated via AIC; $M_{HMM} = 6$. The average duration of browsing states in the HSMM were s_1 : 11.57, s_2 : 9.95, s_3 : 10.16, and s_4 : 11.57, which are longer than the average durations of states in the HMM (s_1 : 1.80, s_2 : 1.00, s_3 : 1.76, s_4 : 1.00, s_5 : 2.50, s_6 : 1.81). The averages are calculated from the sequences of the estimated browsing states. Note that the lengths correspond to the number of gaze actions, i.e., the number of item regions looked at by the participant. The results show that the duration distributions of hidden states of an HSMM contribute to estimate reasonable browsing states since browsing states we want to estimate here correspond to subtasks (see Chapter 2) and are not supposed to change frequently.

Estimation of Browsers' Interest

According to the findings from gaze analysis on choice behavior, visual attention to competing items increases significantly when a consumer is close to making a decision [Shimojo et al., 2003]. Therefore, we assume that the estimation of browser's interest can be improved by identifying when the browser enters the *evaluation* state, and we propose an estimation method of browsers' interest from eye movements, simply by finding items that the browser looked at the most times during *evaluation* states (s_2 and s_4).

We evaluated our proposed model against three methods: HMM, Sliding window, and Baseline. Leave-one-subject-out cross validation was used; that is, gaze data of one participant were used as test data and the remaining data were used to train the model. In the Sliding window method, a sliding window of size 20 is

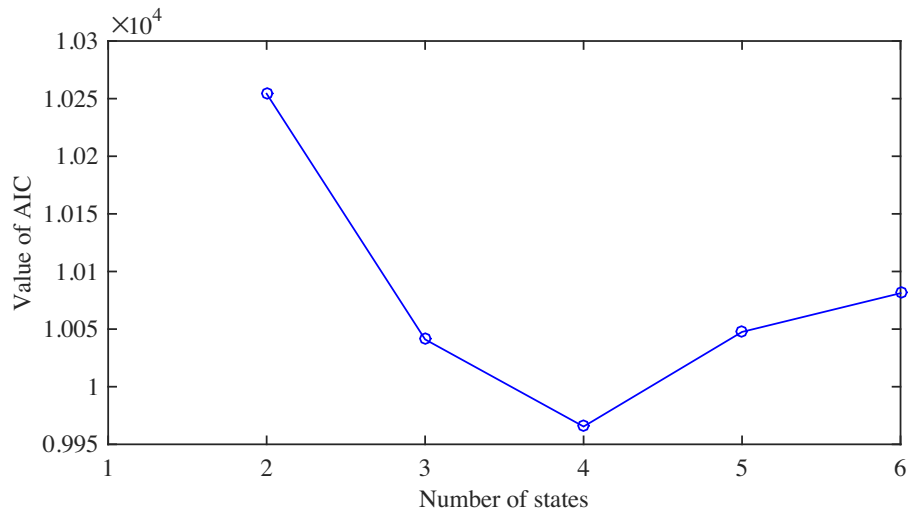


Figure 5.5: The AIC values of different numbers of states.

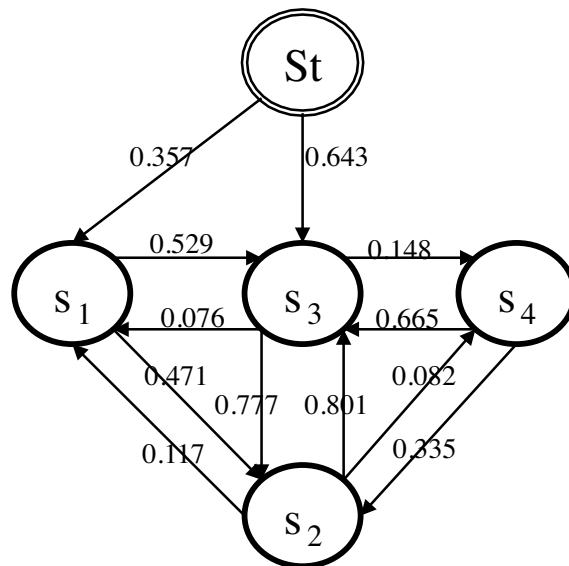


Figure 5.6: The topology of the trained HSMM.

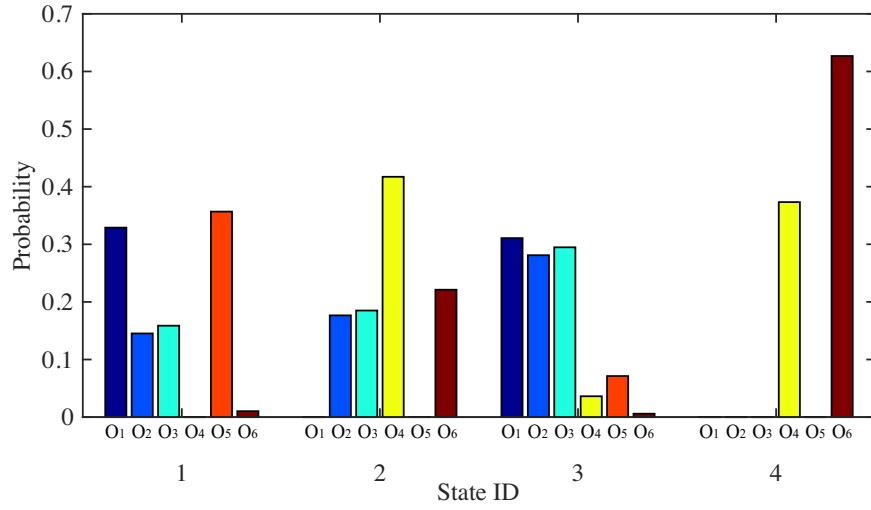


Figure 5.7: Output probabilities of each state in HSMM. The color of the bars corresponds to the color of gaze features in Figure 4.2 (b).

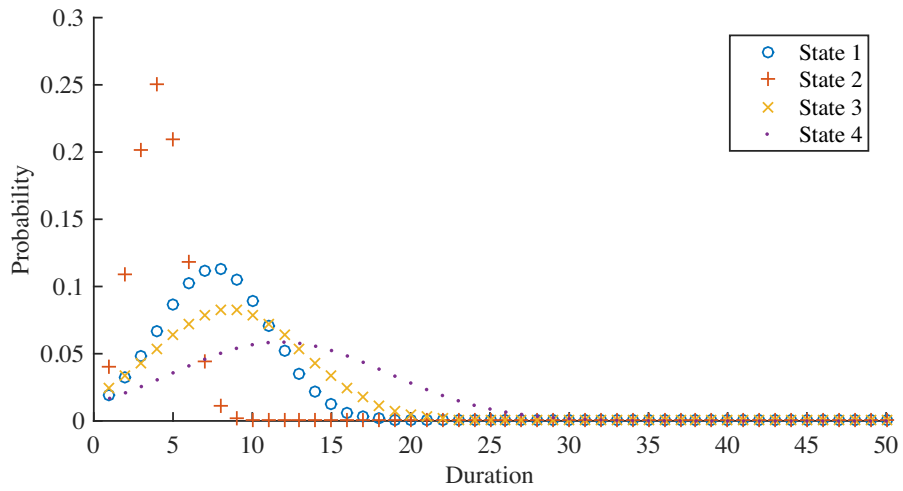


Figure 5.8: Probabilities of duration of each state in HSMM.

5.3. Experiment

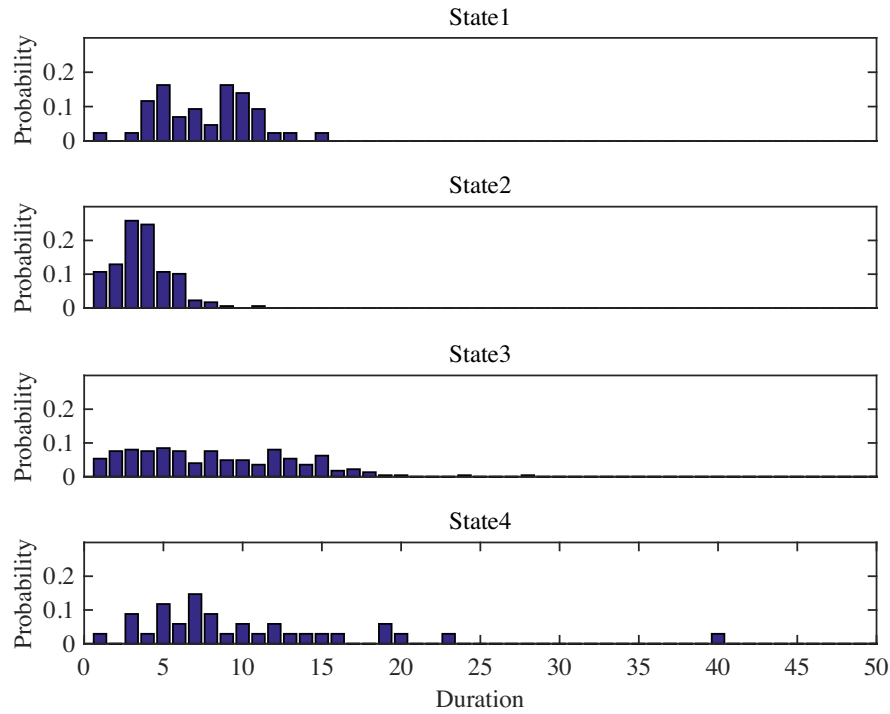


Figure 5.9: Duration distribution of each state before the approximation by a Gaussian distribution.

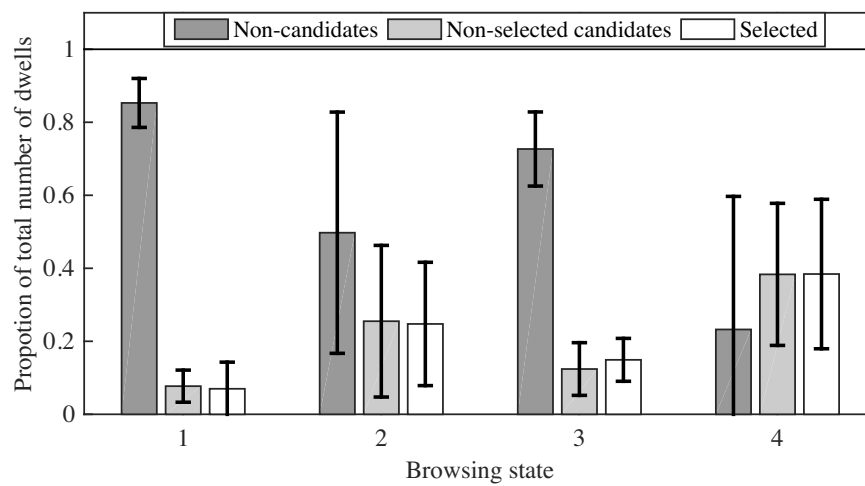


Figure 5.10: The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 2). The whiskers represent standard deviation.

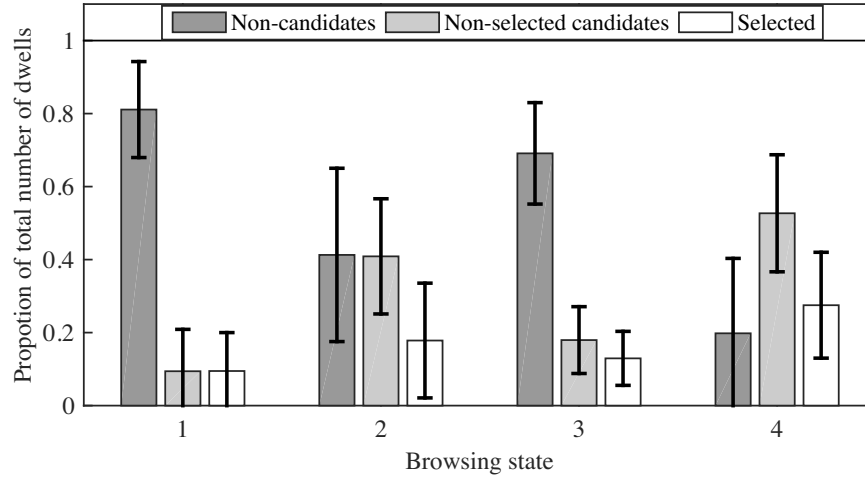


Figure 5.11: The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 3). The whiskers represent standard deviation.

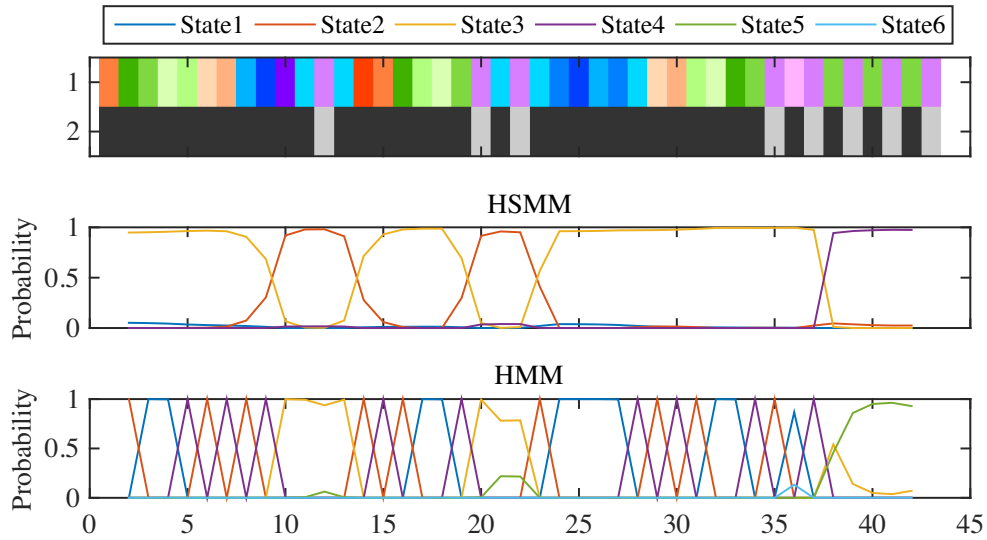


Figure 5.12: An example timeline of browsing states estimated by the trained HSMM and the trained HMM for the gaze region sequence of the 4th trial of the 8th participants (refer to Figure 5.4).

Table 5.3: The estimation of the selected item.

HSMM	HMM	Sliding window	Baseline
0.422	0.469	0.219	0.406

Table 5.4: The estimation of the set of considered items (selected items and candidates).

	Precision	Recall	F1 score
HSMM	0.326	0.763	0.457
HMM	0.358	0.785	0.492
Sliding window	0.298	0.301	0.300

first applied to gaze feature sequences. Then the browsing state at the j -th gaze region, r_j , was determined by finding the nearest distribution in initial values of B to the distribution of gaze actions of interval $[r_{j-9}, r_{j+10}]$. Recall that the initial values of B in both the HMM and the HSMM are the mean points of the M clusters of the distributions of the gaze actions obtained by applying a sliding window of size 20. The distance of distributions here are Euclidean distance.

As with HSMM, *evaluation* states estimated by HMM and Sliding window were identified based on which kind of items were more likely to be looked at in each estimated state. In the Baseline method, the entire gaze sequence was considered when finding the item with the longest duration and was taken to be the estimation of the selected item.

Table 5.3 shows that the use of browsing states estimated by the HMM enables the estimation of selected items with the highest accuracy. In addition, we also conducted the estimation of the set of considered items (selected items and candidates). For this analysis, we only consider gaze sequences of choice tasks with more than two candidates. This is because gaze behavior on selected items and candidates can be very similar in a situation where there are multiple attractive candidates in the catalog. For the estimation of considered items, we find all items looked at during *evaluation* states. Thus, the results of candidate estimation using the Baseline method is omitted. Table 5.4 shows that the use of the HMM achieved the estimation with the highest accuracy.

5.4 Discussion

We evaluated the proposed browsing state model through the eye tracking experiment in a catalog browsing situation. The proposed model estimated browsing

states that have similar characteristics with decision stages defined in the previously proposed consumer decision making model [Russo and Leclerc, 1994]. The results indicate that the proposed model can estimate reasonable browsing states in an unsupervised manner.

We compared two browsing state models based on HSMM and HMM. The results showed that the models have advantages and disadvantages for estimating browsing states. The estimated browsing states by the HMM improved the estimation accuracy of browsers' interest most effectively; however, the states change very frequently, which make it hard to interpret the meanings of the states. Meanwhile, the estimated browsing states by HSMM were able to split choice process into several phases, which matches the findings from existing studies of choice process. Thus, we concluded that the use of HSMM is more suitable to estimate interpretable browsing states.

Analysis Methods of Estimated Browsing States

The estimated browsing states in this chapter can be used for detailed analysis of content browsing behavior. For the analysis of multi-alternative choice behavior, the duration and transitions of the estimated browsing states can be used as one of measurements to assess the choice process. For example, the duration of browsing states can be interpreted as how long a browser takes to eliminate alternatives, which directly characterize important aspects of the choice process such as decision strategy.

Individual Differences among Browsers

There could be individual differences among browsers; for example, in a multi-alternative choice situation, some browser may take more time to evaluate alternatives than others. One possible solution to capture such individual differences is to build multiple models based on user profiling. For this, we would need to build a data set that captures details about individual characteristics of browsers and apply cluster analysis.

Chapter 6

Model Adaptation for Choice Behavior Analysis

6.1 Introduction

In Chapter 2 through Chapter 5, we presented the proposed framework for semantic interpretation of eye movements. In Chapter 2, the overview of the proposed framework is presented. In Chapter 4, we proposed the method to encode eye movements into gaze actions based on the designed structure defined in Chapter 3. In Chapter 5, the browsing state model is trained in a unsupervised manner; that is, the characteristics and number of browsing states and their transitions in a certain decision making session are estimated simultaneously. The unsupervised approach can be used even when we do not have any prior knowledge on browsing states; meanwhile, when we focus on a specific content-browsing-domain, and when the knowledge of the domain is well accumulated by existing studies, it is also possible to adapt the browsing state model using the prior knowledge. In this chapter, we describe a method for adapting the proposed framework to analyze a specific content browsing situation: choice behavior in catalog browsing.

Various aspects of choice behavior have been investigated using eye tracking data, including viewers' preference [Sugano et al., 2014, Shimojo et al., 2003], decision tasks [Pfeiffer et al., 2014], and motivations/time pressure [Pieters and Warlop, 1999]. The advantage of tracking eye movements is that it enables the detailed analysis of decision processes, which includes the temporal changes of the decision makers' cognitive states. As mentioned in Chapter 4, several studies have shown that the decision process

consists of a few different phases [Orquin and Loose, 2013]; that is, decision makers split their decision making task into smaller subtasks such as the screening of alternatives. These phases are sometimes referred to as *decision stages* [Russo and Leclerc, 1994, Gidlöf et al., 2013] or *attention phases* [Orquin and Loose, 2013]. In this thesis, we use the term “decision stages”. The existence of decision stages has been confirmed in many works [Russo and Leclerc, 1994, Wedell and Senter, 1997, Clement, 2007, Glaholt and Reingold, 2011, Reutskaja et al., 2011, Glöckner and Herbold, 2011, Gidlöf et al., 2013].

In early versions of decision stage models, before eye tracking was widely used, the decision process was considered to be roughly separated into two types of decision stages: *screening* and *evaluation* [Payne, 1976]. In the screening stage, decision makers obtain an overview of the information about all products; then, in the evaluation stage, they evaluate a few remaining alternatives to make a selection. To our knowledge, Russo et al. were the first to utilize eye tracking to model decision stages [Russo and Rosen, 1975, Russo and Leclerc, 1994]. Russo measured participants’ eye movements in a supermarket-like experimental environment where real products were displayed on a shelf. Through the experiment, Russo et al. confirmed the existence of *verification* stage after the above-mentioned two decision stages. Their work led to a lot of follow-up work on decision stages in decision making [Wedell and Senter, 1997, Clement, 2007, Glaholt and Reingold, 2011, Reutskaja et al., 2011, Glöckner and Herbold, 2011, Gidlöf et al., 2013].

Many questions still remain in decision stage analysis; a fundamental problem is that there is no concrete method for the identification of decision stages. Existing methods employ a heuristic approach in which top-down rules are given by analysts based on their observations. This caused disagreement about the number of decision stages and their interpretations among previous studies. Some studies [Gidlöf et al., 2013, Russo and Leclerc, 1994] noted the decision stages cannot be clearly separated like the previous studies assumed. The boundary between stages is further blurred due to the occasional interruption into other stages. For example, decision makers often go back to the screening stage to re-input item information during the evaluation stage. Thus, a concrete theory on the quantity and type of decision stages is not established, and this may change with task characteristics. Accordingly, it is not feasible to model the decision stage interruptions by defining top-down rules that cover all possible cases.

In this chapter, we adapt our browsing state model proposed in Chapter 5 to trace temporal changes of the decision states. That is, we incorporate the browsing state model with some top-down assumptions based on findings from studies on decision stage analysis. The estimated browsing states using the adapted model correspond to decision stages in choice behavior analysis. More specifically, this chapter consists of two steps: (1) building of a choice behavior model to define and identify browsing states and (2) supervised learning of the browsing state model (proposed in Chapter 5) to estimate browsing states only with gaze actions. At step (1), we define browsing states in choice behavior based on a strong but domain-specific indicator: how frequently the selected item is looked at by browsers. At step (2), the browsing state model is learned via a supervised manner using the identified browsing states. Once the browsing state model is learned, browsing states of a newly observed eye tracking data can be estimated without the information about selected items.

In the following sections, we introduce related studies on choice behavior analysis in Section 6.2. Then, the definition of browsing states and its evaluation are presented in Section 6.3. Finally, the estimation method based on the browsing state model is presented in Section 6.4.

6.2 Choice Behavior Analysis

This section first presents more details on previously proposed decision stage models and their limitations [Russo and Leclerc, 1994, Gidlöf et al., 2013]. Second, we present several existing information-search gaze models that take a probabilistic approach. Third, we introduce findings on content browsing behavior in different task complexities since we use the task complexities to evaluate the proposed approach. Finally, we present issues and the approach of this study to analyze choice behavior.

Decision Stages in Choice Process

As mentioned above, Russo et al. pioneered decision stage analysis using eye tracking [Russo and Leclerc, 1994]. They investigated the existence of decision stages in choice behavior and their characteristics. As a result of their eye tracking experiments, they found that the decision processes consist of three stages: *orientation*, *evaluation*, and *verification*. In their paper, each stage is identified based on a sequence of “dwells” on alternative items. Recall that a dwell is a set of consec-

utive fixations on each alternative ¹. Dwell-sequence analysis is the commonly used approach for process tracing of decision stages [Glaholt and Reingold, 2011, Orquin and Loose, 2013]. The orientation stage is defined to be the stage that occurs before the first “re-dwell” (a dwell on an already-examined item) appears. The evaluation stage is between the first and the last re-dwell and the verification stage is after the last re-dwell. This study leads to a lot of follow-up work on decision stage analysis; however, the validity of Russo’s work has been questioned by some researchers due to the fact that the researchers were manually tracking participants’ eye movements through a one-way mirror [Pfeiffer et al., 2014].

More recently, Gidlöf et al. re-evaluated Russo’s three-stage model, calling it the Natural Decision Segmentation Model (NDSM), and used modern eye tracking technology [Gidlöf et al., 2013]. In NDSM, the information of the selected item is used as a stage divider, i.e., the evaluation stage in NDSM is defined to be between the first and the last dwell on the selected item. Their model was evaluated through a real-world super market experiment by measuring how the identified decision stages can describe differences between two tasks (search and decision) compared to Russo’s model. Their research established a difference in browsing behavior between the two tasks; however, their method of identifying different decision stages still had some limitations. As noted in their paper, the stages can be often interrupted by other stages, for example, going back to the orientation stage to re-input item information during the evaluation stage. Instead of defining rules to cover the interruption of decision stages, Gidlöf et al. assumed each stage is dominated by a specific function of the decision process even if it contains elements of other stages.

Probabilistic Approach in Modeling Information Processing States

There are a few probabilistic gaze models in consumer decision process. Stüttgen et al. proposed a satisfactory decision process model using eye tracking data [Stüttgen et al., 2012]. The satisfactory decision process model is a decision making model that assumes decision makers choose alternatives that satisfy a set of specific criteria. In their model, information search behavior is represented by HMM which has two states: local and global search state based on the above mentioned Liechty’s local/global attention model [Liechty et al., 2003]. Shi et al.

¹The term of ‘fixation’ here indicates a visual fixation between saccades. The term is also used in Russo’s paper [Russo and Leclerc, 1994]; however, Russo’s “fixation” corresponds to a glance or a dwell in this thesis.

proposed a hierarchical HMM to uncover information acquisition process on catalogs with attribute-by-product matrix layouts [Shi et al., 2013]. Their model consists of two layers of hidden states, in which the lower-layer hidden states correspond to types of *information acquisition processes*, and the higher-layer hidden states correspond to *strategy*. For their model observations, they employ probabilities of three of the types of transitions among AOIs (regions that describe an attribute of a product) on a screen: attribute-based (a transition between regions of the same attribute but different items), product-based (a transition between regions of different attributes but for the same item), and others. Although their model can represent similar states to decision stages, it can be applied only to eye movements for catalogs with attribute-by-product matrix layouts. In most situations where the differences between attribute-based and product-based information search behavior are unclear, we need to reconsider which features of gaze should be analyzed.

Browsing States and Task Complexities

Information processing varies with the parameters of decision making tasks [Payne, 1976]. When people are faced with a decision between two alternatives, they employ a compensatory decision process in which all available information of alternatives is considered together to make a comprehensive decision. In contrast, when people face with a more difficult decision task (with more alternatives), they tend to employ a non-compensatory decision process in which some of the available alternatives are eliminated quickly based on some specific criteria (e.g., filtering). After narrowing the alternatives down, people finally evaluate the alternatives using a compensatory decision process. This is simply because of a limited information processing capability of decision makers. Note that the compensatory decision process corresponds to the evaluation state in our model.

Issues and Approach of This Study

As discussed above, there are two main issues when modeling temporal changes of browsing states that correspond to decision stages in choice behavior: (1) managing the diversity of browsing states and (2) characterizing features of gaze for identifying the browsing states. To address issue (1), this work provides a bottom-up framework in which probabilistic relations between decision stage and gaze

behavior are learned from recorded gaze data. A HMM was employed to represent bi-directional transitions among browsing states. The HMM is similar to the browsing state model proposed in Chapter 5; however we employ some top-down assumptions about the number and types of hidden states. To address issue (2), we employ information about when the selected items are being looked at as a strong indicator of browsing states in choice behavior. That is, we assume that how often the selected item is looked at will change depending on the browsing state and employ interval lengths between dwells on the selected item.

Our approach is evaluated by measuring the description capability of differences among tasks. Here, we assume that the number of remaining alternatives (after the elimination process) affects the temporal duration of the evaluation state. Ideally, if our assumption is correct and if only one alternative remains after screening, the evaluation state would not be observed. Meanwhile, if multiple alternatives (strong candidates) remain after screening, the duration of the evaluation state would increase.

6.3 Defining Browsing States in Choice Behavior

This section describes the details of the proposed choice behavior model and the results of its evaluation. First, basic assumptions about the browsing states are described. Next, we explain how we formulate the assumptions for the probabilistic model. Finally, our browsing model is evaluated using eye tracking data collected in Chapter 5.

6.3.1 Basic Assumptions of Browsing States

Here we make a couple of assumptions of browsing states in multi-alternative choice situations. First, we assume the choice behavior can be described by two types of browsing states: *exploration* and *evaluation*. The exploration state is a decision state where decision makers aim to obtain broad information of items to screen items out or to confirm their decision. Meanwhile, the evaluation state is a decision state where a decision maker aims to obtain detailed information of fewer items to evaluate them as strong candidates. We define the correspondence between our browsing states and the three decision stages in the previously proposed three stage model [Russo and Leclerc, 1994, Gidlöf et al., 2013] as follows: the exploration state corresponds to the orientation and the verification stage, and

the evaluation state corresponds to the the evaluation stage. Although the orientation stages and the verification stages are qualitatively distinguished by Russo et al., the gaze behavior for each state would be expected to be similar since the two stages are identified by applying the same process from the beginning of the gaze sequence forward or from the end to the beginning.

Second, we assume that the browsing states can be characterized/identified by examining how frequently dwells on the selected item occur. That is, decision makers are expected to look at the selected item more often when they are in the evaluation state than when they are in the exploration state. In previous decision stage models, re-dwells (a dwell on an item that is already looked at more than once) [Russo and Leclerc, 1994] or the first and the last dwells on the selected item [Gidlöf et al., 2013] are considered an important cue to identify decision stages. We integrated those cues, in particular, our method considers how long it takes until a re-dwell on the selected item occurs in each browsing state (interval lengths between dwells on the selected item).

6.3.2 The Proposed Choice Behavior Model

Suppose a catalog contains information of a set of items $\mathcal{I}_{\text{All}} = \{1, \dots, N\}$, and a decision maker selects an item out of the item set. Eye tracking data of the decision maker are first represented as a sequence of dwells on items $R = (r_1, \dots, r_J)$ ($r_j \neq r_{j+1}, r_j \in \mathcal{I}_{\text{All}}$), where J is the total number of dwells in the decision making session. Let us denote item labels as I_S and I_{NS} that correspond to the selected item and the other items, respectively. Here, we introduce a function $\text{Label} : \mathcal{I}_{\text{All}} \rightarrow \{I_S, I_{NS}\}$, where $\text{Label}(r_j)$ indicates the item label of the j -th dwell, r_j . Assuming the selected item is looked at K times in the sequence, a sequence of timings of dwells on the selected items is obtained as $J^S = (j_1^S, \dots, j_K^S)$ ($j_k^S \in \{1, \dots, J\}, j_k^S < j_{k+1}^S$) by finding all timings that satisfy $\text{Label}(r_{j_k^S}) = I_S$.

The dwell sequence, R , is then segmented using the timings of dwells on the selected item, J^S , as $\tau_k = [j_k^S + 1, j_{k+1}^S]$ ($k \in \{1, \dots, K-1\}$) (see Figure 6.1). At the end of the dwell sequence, the decision maker makes a choice; and thus, the attention of the decision maker just after the last dwell, r_J , is considered to be on the selected item. Thus, when the last dwell, r_J , is not on the selected item (i.e., $j_K^S \neq J$), we consider an additional segment $\tau_K = [j_K^S + 1, J + 1]$ after τ_{K-1} . Meanwhile, the segment before the first dwell on the selected item $\tau_0 = [1, j_1^S]$ should be treated differently to the other segments since the timing of the first dwell on the selected item depends on exogenous factors of content layout rather than brows-

ing states of decision makers. For this, we consider *start state* in addition to the above mentioned two browsing states.

Let us denote the two browsing states, the exploration and the evaluation state, as S_{exp} and S_{eva} , respectively, and the start state as S_{st} . A sequence of segments is defined as $\tau = (\tau_0, \tau_1, \dots, \tau_{K'})$, where $K' = K + 1$ when $j_K^S \neq J$, and $K' = K$ when $j_K^S = J$. Assuming the browsing states do not change during the segments, what we want to estimate here is a sequence of browsing states in each segment $Q = (q_0, q_1, \dots, q_{K'})$ ($q_k \in \{S_{\text{exp}}, S_{\text{eva}}\}, k \neq 0$) using the length of the segments as a cue. Note that every decision process starts with the start state in our model (i.e., $q_0 = S_{\text{st}}$). In particular, we formulated how long it takes until a re-dwell on the selected item occurs (i.e., lengths of the segments) in each browsing state as follows.

Exploration State

The exploration state is defined to be a state where a decision maker aims to explore the information of an entire catalog. Thus, each item in the catalog is considered to be looked at with equal frequency regardless of whether or not it is the selected item. Assume that the k -th segment is the exploration state and contains the j -th dwell and the $(j - 1)$ -th dwell. Since the $(j - 1)$ -th dwell is not the end of the segment, the $(j - 1)$ -th dwell is on a non-selected. As mentioned above, the item on the $(j - 1)$ -th dwell and the j -th dwell is always unequal $r_{j-1} \neq r_j$. Thus, when the $(j - 1)$ -th dwell is on a certain item, the j -th dwell can only be on one of the $(N - 1)$ items, where N is the total number of items on the catalog as mentioned above. Accordingly, the probabilities of the j -th dwell being on a non-selected item and on the selected item are represented as $P(\text{Label}(r_j) = I_{\text{NS}} | \text{Label}(r_{j-1}) = I_{\text{NS}}, q_k = S_{\text{exp}}) = (N - 2)/(N - 1)$, and $P(\text{Label}(r_j) = I_{\text{S}} | \text{Label}(r_{j-1}) = I_{\text{NS}}, q_k = S_{\text{exp}}) = 1/(N - 1)$, respectively.

The probability of the k -th interval lasting l dwells is described as a geometric distribution

$$P(|\tau_k| = l | q_k = S_{\text{exp}}) = \left(\frac{N - 2}{N - 1} \right)^{(l-2)} \frac{1}{N - 1}. \quad (6.1)$$

Evaluation State

The evaluation state is defined to be a state where a decision maker aims to obtain more detailed information of a set of specific items (strong candidates). Therefore,

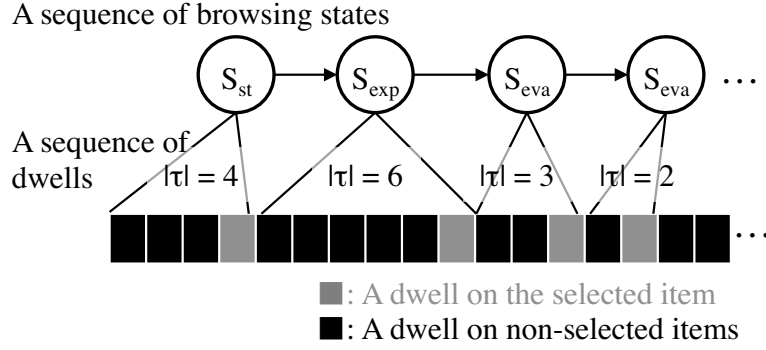


Figure 6.1: Segments of browsing states.

in the evaluation state, the probability of a dwell on the selected item is considered to be higher than one in the exploration state. Since we do not know the probability, we introduced a parameter p_{eva} to represent the probability of a dwell on the selected item in the evaluation state; that is, $P(\text{Label}(r_j) = I_S | \text{Label}(r_{j-1}) = I_{NS}, q_j = S_{eva}) = p_{eva}$. Thus, the probability of the j -th dwell being on a non-selected item is represented as $P(\text{Label}(r_j) = I_{NS} | \text{Label}(r_{j-1}) = I_{NS}, q_j = S_{eva}) = 1 - p_{eva}$. Here, the value of p_{eva} is estimated from eye tracking data in this study. Note that the higher value of estimated p_{eva} indicates that the selected item is more likely to be looked at in the evaluation state.

The probability of the j -th interval lasting l dwells is described as a geometric distribution

$$P(|\tau_k| = l | q_k = S_{eva}) = (1 - p_{eva})^{(l-2)} p_{eva}. \quad (6.2)$$

A Hidden Markov Based Choice Behavior Model

Since we want to represent free transitions of the browsing states, we employ a HMM. The proposed HMM is denoted as $\lambda_{\text{choice}} = (\mathcal{S}, \pi, A, \mathcal{L}, C)$. $\mathcal{S}$ is a set of browsing states; $\mathcal{S} = \{S_{st}, S_{exp}, S_{eva}\}$. π is a set of initial state probabilities; $\pi = (\pi_{st}, \pi_{exp}, \pi_{eva}) = (1, 0, 0)$ since our model assumes decision process always starts from the start state, S_{st} . A is a 3×3 state transition probability matrix, where $a_{i,j}$ indicates the transition probability from the i -th state to the j -th state. In our model, $a_{i,j}$ is always set to zero if the j -th state corresponds to the start state, S_{st} . That is, no transition from any states to the start state, S_{st} , occurs including self-transition.

Possible observations of the model are the lengths of the segmented intervals,

τ ; $\mathcal{L} = \{1, \dots, L_{\max}\}$, where $L_{\max}$ is the maximum length of segmented intervals. C is a set of output probabilities; $C = \{c_i(l)\}_{i \in S, l \in \mathcal{L}}$, where $c_i(l)$ indicates the probability of the segment of length l at the i -th state. Note the probability of an output $l \in \mathcal{L}$ in the exploration state and the evaluation state, $c_{S_{\text{exp}}}(l) = P(|\tau_k| = l | q_k = S_{\text{exp}})$ and $c_{S_{\text{eva}}}(l) = P(|\tau_k| = l | q_k = S_{\text{eva}})$, are defined as the Equation (1) and the Equation (2), respectively. The output probabilities in the start state are denoted as $c_{S_{\text{st}}}(l) = P(|\tau_0| = l)$. The output probabilities are normalized so that the sum of output probabilities in each state is 1.

Estimation of the Model Parameters and Identification of Browsing States

The unknown parameters to be estimated here are p_{eva} , the probability of dwells on the selected item in the evaluate state, $\{c_{S_{\text{st}}}(l)\}_l$, the output probabilities of the start state, and A , the transition probabilities of hidden states. Since the first output of the sequence is defined to be from the start state, the output probabilities of S_{st} are simply calculated as the frequency distribution of the first outputs of each sequence in training gaze data. The rest of model parameters are estimated by the Baum-Welch algorithm [Baum et al., 1970]. Once the model parameters are estimated, the sequence of browsing states, Q , can be estimated from the sequence of segments, τ , using the Viterbi algorithm [Viterbi, 1967].

6.3.3 Evaluation

Our catalog browsing model is evaluated by comparing with two previously proposed decision process models: Russo's three stage model (R&L) [Russo and Leclerc, 1994] and the Natural Decision Segmentation Model (NDSM) [Gidlöf et al., 2013] using eye tracking data collected in Chapter 5. In the experiment, the task criterion specifies the number of candidates among a set of items on a screen. Then, our approach is evaluated by measuring how the model can segment reasonable intervals for browsing states in terms of (1) description capability of differences among the tasks and (2) accuracy in identifying when a participant is exhibiting comparison behavior among candidate items.

Estimated Model Parameters

The estimated state transition probabilities using the data of all participants are shown in Figure 6.2. The result shows that the transition from the start state to the exploration state is more likely than one from the start state to the evaluation

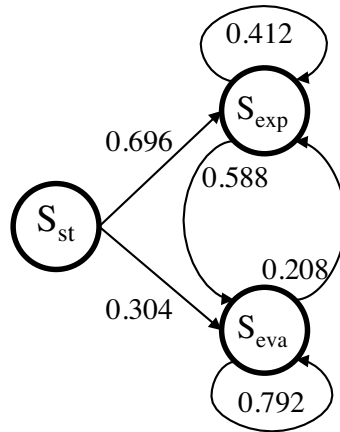


Figure 6.2: The estimated transition probabilities of the browsing states.

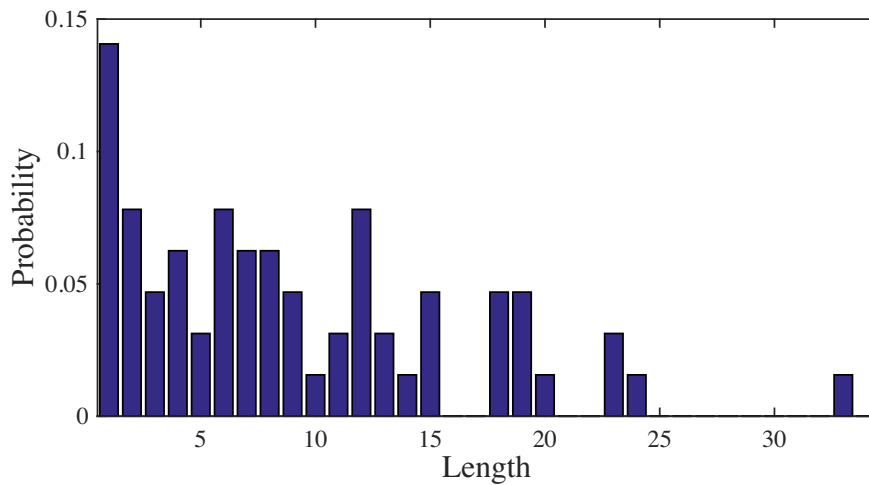


Figure 6.3: The output probabilities of the start state, S_{st} , estimated using eye tracking data of all participants.

state. Moreover, the transition from the exploration state to the evaluation state is of higher probability than the opposite transition. This indicates that decision makers tend to start from the exploration state and then gradually shift to the evaluation state.

Figure 6.3 shows the output probabilities of the start state. Figure 6.4 shows the output probabilities of the exploration/evaluation states and the frequency distribution of lengths of all segments except the first segment (i.e., $\tau_1, \dots, \tau_{K'}$). The probability of dwells on the selected items in the evaluation state was estimated as $p_{eva} = 0.544$. This result shows that decision makers are more likely to look at the selected items within short intervals in the evaluation state.

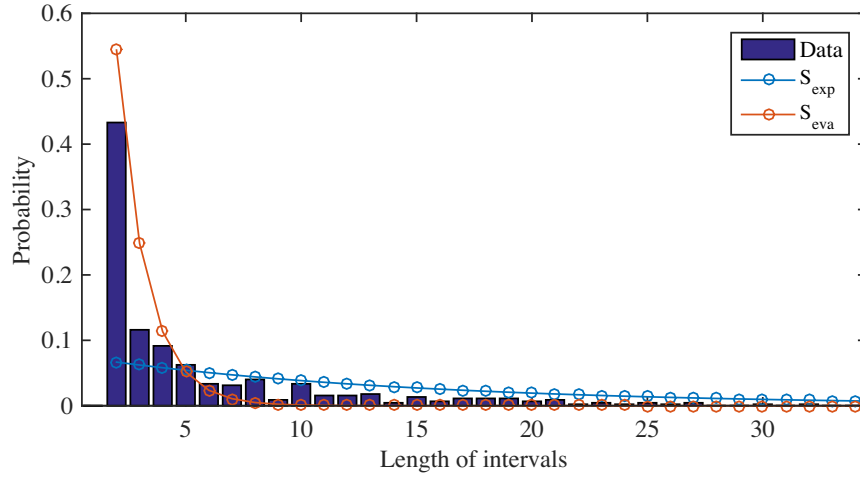


Figure 6.4: The output probabilities of the exploration state (S_{exp}) and the evaluation state, S_{eva} , estimated using eye tracking data of all participants. The bars represent the frequent distribution of segment lengths.

Example Sequences of Identified States

Figure 6.5 shows four example sequences of identified browsing states by the proposed model and the other two comparison models. The comparison models identify browsing states similarly to each other only when the first dwell on the selected item and the first re-dwell occurs at the same time (see the bottom-left example). However, in most other cases, participants continued examining other items even after looking at the selected item, such as in the top-left example. Moreover, the interruptions of browsing states cannot be identified by the comparison models. In the bottom-right example, there was a gap from the 12th dwell to the 21st dwell where the selected item is less frequently looked at; however, R&L and NDSM both identify the evaluation state so as to occupy the whole sequence, and the gap is identified only with the proposed method.

Difficulty of Decision Tasks and Browsing States

Figure 6.6 shows the mean ratio of the two browsing states with each task type. The proposed method and R&L show similar effects of the task, where the ratio of the evaluation states gets higher as the number of strong candidates (task difficulties) increases. There was a significant difference of the ratio of the evaluation state between the task 1 and the task 3 using R&L model (Welch's t-test; $p = .037$). Although we did not observe any significant difference for the results

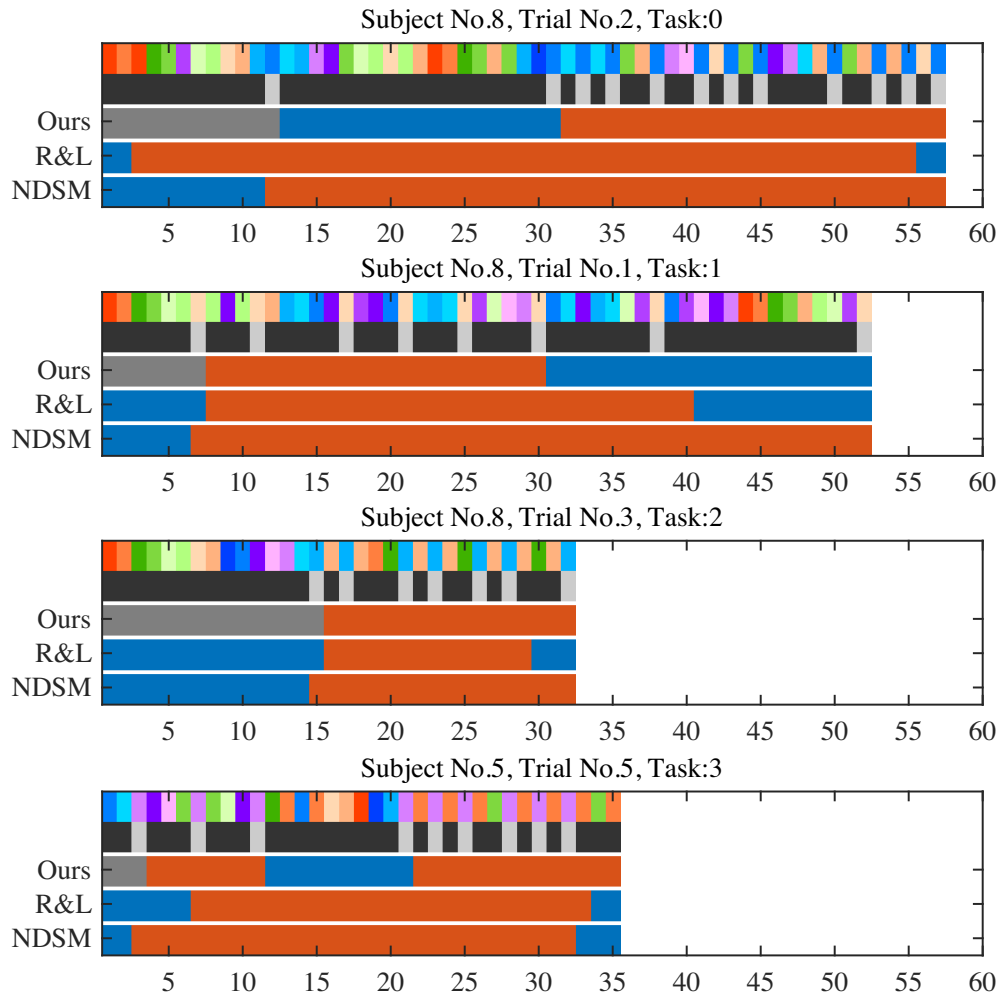


Figure 6.5: Examples of estimated browsing state sequences. For each sequence, the color in the top row corresponds to the color of regions in Figure 5.3. In the second row, the timings of the dwells on the selected item are shown by light gray highlights. From the third row to the bottom row, the blue and red intervals correspond to the exploration state and the evaluation state, respectively.

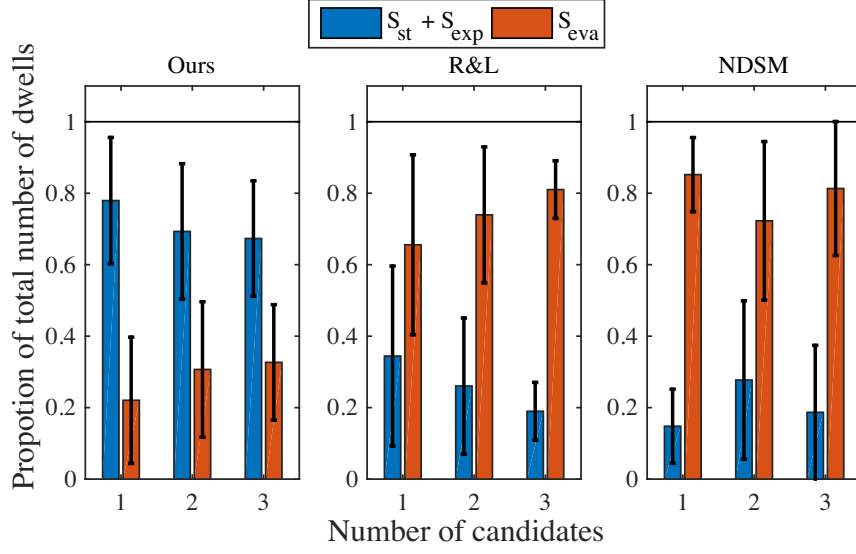


Figure 6.6: The proportion of dwells in each browsing state. The whiskers represent standard deviation.

of the proposed model and NDSM, the proposed model shows the similar effects of the task to ones with R&L model. These results indicate that our model and R&L can describe differences of browsing behavior between different tasks when compared with NDSM. Note here that we did not consider eye tracking data in tasks where the number of candidates is 0 since the difficulty of the decision task is hard to interpret.

The ratio of the evaluation and exploration states identified using our model is biased towards the exploration state more than the other two comparison models. This indicates that our model is more sensitive to the interruptions of the exploration/evaluation states, which often appear as shown in Figure 6.5. This result shows that allowing bi-directional transitions of browsing states in the state model is more effective when modeling browsing states.

Dwells on Non-selected Candidates and Browsing States

Recall that the candidates are the set of items that satisfy the task (see Section 5.3 for the details). It is important to consider how users regard both selected and non-selected items to understand the characteristics of browsing states in tasks with more than two candidates. Figure 6.7 and Figure 6.8 show the proportion of dwells on each item type: the selected item, non-selected candidates, and others. Figure 6.7 shows the results in tasks with two candidates, and Figure 6.8 shows the results in tasks with three candidates. In both figures, non-selected candidates

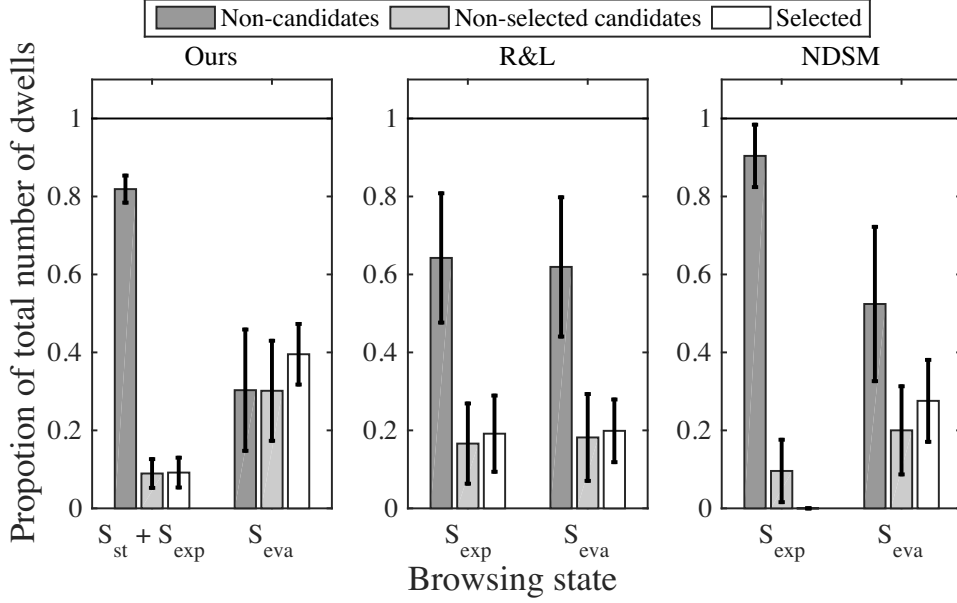


Figure 6.7: The proportion of dwells on selected items, non-selected candidates, and the other items when the number of candidates is 2. The whiskers represent standard deviation.

are more likely to be looked at in the evaluation states, S_{eva} , compared to the exploration states, S_{exp} . There were significant differences of the ratio of the dwells on non-selected candidates between the exploration state and the evaluation state using our model (Welch’s t-test; $p < .001$ for two-candidate-tasks and $p = .022$ for three-candidate-tasks). Using NDSM, there was a significant difference between the exploration state and the evaluation state in two-candidate-tasks (Welch’s t-test; $p = .009$). Using R&L, we did not observe any significant difference. Only when our model was used, the significant differences were observed in the both tasks.

6.4 Supervised Estimation of Browsing States

In this section, we propose an estimation method of browsing states that does not require information of selected items. To achieve this, we extend the choice behavior model, λ_{choice} , based on the browsing state model, λ_{HSMM} , proposed in Chapter 5. That is, instead of using the information about when the selected items are being looked at, we employ gaze actions proposed in Chapter 4 as a cue to estimate browsing states. The difference between the estimation method of browsing states in this chapter and in Chapter 5 is the learning method. The

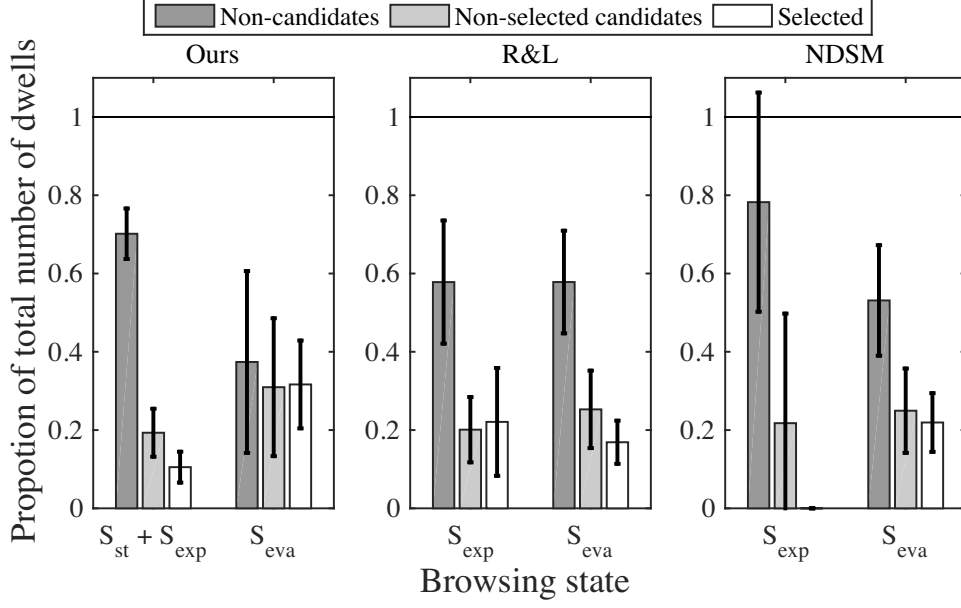


Figure 6.8: The proportion of dwells on selected items, non-selected candidates, and other items when the number of candidates is 3. The whiskers represent standard deviation.

browsing states are estimated in a unsupervised manner in Chapter 5 without any assumptions/content-specific domains; meanwhile, the browsing states in this chapter are estimated in a supervised manner using a training data set of sequences of gaze actions and corresponding browsing states identified by λ_{choice} .

The following sections first present an overview of the proposed estimation method. Next, the details of how to encode eye movements into gaze actions and how our choice behavior model is associated with the gaze actions is presented.

6.4.1 The Overview of The Proposed Estimation Method

As with Chapter 4 and Chapter 5, eye movements are first encoded into a sequence of gaze actions using designed structure. Here, we focus on tri-gram patterns described in 4.2.2. For the later analysis, we assume sequences of gaze actions are obtained as $X_n = (x_2, \dots, x_{J_n-1})$, ($x_j \in \mathcal{O} = \{o_1, \dots, o_6\}$) from a dwell sequence $R_n = (r_1, \dots, r_{J_n})$, where the gaze actions, $\mathcal{O}$, correspond to the patterns as shown in Figure 4.2 (b).

Second, the probabilistic relations between gaze actions and browsing states are learned. Suppose we have a set of gaze-action sequences $\mathcal{X} = \{X_n\}_n$ and corresponding sequences of browsing states $\mathcal{Q} = \{Q_n = (q_2, \dots, q_{J_n-1})\}_n$ obtained

by our choice behavior model, λ_{choice} , as training data. In this situation, the goal is to estimate browsing states from a newly observed gaze data without the information of the selected item. Once the model is learned, we can estimate browsing states of newly observed eye tracking data using the learned model.

6.4.2 Modeling Probabilistic Relations of Gaze Actions and Browsing States

In the previous section, our choice behavior model is denoted as $\lambda_{\text{choice}} = (\mathcal{S}, \pi, A, \mathcal{L}, C)$. To associate gaze actions with browsing states, we here introduce the browsing state model proposed in Chapter 5, $\lambda_{\text{HSMM}} = (\mathcal{S}_{\text{HSMM}}, \mathcal{O}_{\text{HSMM}}, A_{\text{HSMM}}, B_{\text{HSMM}}, P_{\text{HSMM}}, \pi_{\text{HSMM}})$. Then, we modify the browsing state model, λ_{HSMM} , so that it corresponds to the choice behavior model, λ_{choice} .

The modified browsing state model is denoted as $\lambda'_{\text{HSMM}} = (\mathcal{S}'_{\text{HSMM}}, \mathcal{O}'_{\text{HSMM}}, A'_{\text{HSMM}}, B'_{\text{HSMM}}, P'_{\text{HSMM}}, \pi'_{\text{HSMM}})$. A set of hidden states $\mathcal{S}'_{\text{HSMM}}$, initial state probabilities π'_{HSMM} , and state transition probabilities A'_{HSMM} are common to the choice behavior model, λ_{choice} ; that is, $\mathcal{S}'_{\text{HSMM}} = \mathcal{S} = \{S_{\text{st}}, S_{\text{exp}}, S_{\text{eva}}\}$, $A'_{\text{HSMM}} = A$, and $\pi'_{\text{HSMM}} = \pi = (\pi_{\text{st}}, \pi_{\text{exp}}, \pi_{\text{eva}}) = (1, 0, 0)$. $\mathcal{O}'_{\text{HSMM}}$ is a set of gaze action labels; $\mathcal{O}'_{\text{HSMM}} = \mathcal{O} = \{o_1, \dots, o_6\}$ (see Figure 4.2 (b)). B'_{HSMM} is a 3×6 output probability matrix, where $b'_{i,m}$ indicates the output probability of the m -th gaze action at the i -th browsing state.

P'_{HSMM} is a $3 \times L_{\text{max}}$ duration distribution matrix, where $p'_{i,l}$ indicates the probability that the i -th state last for l -time units. The duration distribution matrix, P'_{HSMM} , represents interval lengths between dwells on the selected item in the choice model, λ_{choice} . That is, $p_{i,l} = c_i(l)$, where $c_i(l)$ indicates the probability of the segment of length l at the i -th state in the choice model (see Section 6.3.2 for the details). Using a training gaze dataset, a set of gaze action sequences $\mathcal{X}$ and corresponding browsing state sequences $\mathcal{Q}$, the unknown output probabilities, B'_{HSMM} , are simply estimated as follows:

$$b'_{i,m} = \frac{\sum_n \sum_j \mathbf{I}_{q_j=i} \cdot \mathbf{I}_{x_j=o_m}}{\sum_n |X_n|}, \quad (6.3)$$

where $\mathbf{I}_A$ is an indicator function that has the value 1 when A is true and the value 0 otherwise. Once the parameters are obtained, the browsing states of a newly observed gaze data can be estimated by maximum a posterior estimation.

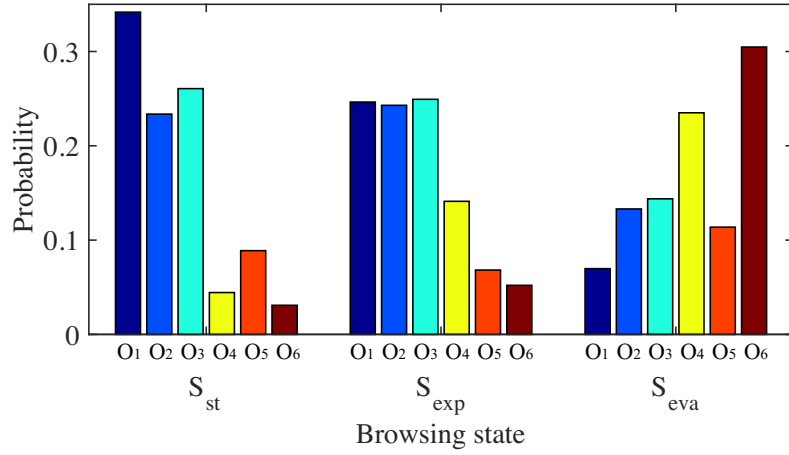


Figure 6.9: The output probabilities of each browsing state. The color of the bars corresponds to the color of gaze actions in Figure 4.2 (b).

6.4.3 Evaluation

The proposed estimation method is evaluated using gaze data collected in Section 5.3. The output probabilities of the HSMM estimated by the Equation 6.3 with gaze data from all participants are shown in Figure 6.9. The results show that the output probabilities of the start state and ones of the exploration state are more similar compared to ones of the evaluation state. In the evaluation state, the gaze action that consists of contrast relations (o_4) and comparison patterns (o_5 and o_6) occur more frequently than in the start state and the exploration state. This indicates that decision makers tend to browse catalogs more freely by ignoring groups of items in the evaluation state.

Consistency with the Identified Browsing States

Subject-based leave-one-out cross validation was used; that is, gaze data of one participant was used as test data and the remaining of the data was used to train the model. We compare our proposed method (HSMM) against two methods: hidden Markov model based estimation (HMM) and maximum likelihood estimation (ML). In the HMM method, a three state HMM ($S_{HMM}, O_{HMM}, A_{HMM}, B_{HMM}, \pi_{HMM}$), where $S_{HMM} = S$, $O_{HMM} = O$, $\pi_{HMM} = \pi$ is trained with the training gaze data: a set of gaze action sequences, $\mathcal{X}$, and corresponding browsing state sequences, $\mathcal{Q}$. Then, the browsing states can be estimated by the Viterbi algorithm using the trained HMM. In the ML method, the

Table 6.1: The consistency with the ground truth. The top row indicates the consistency when we consider that the exploration state S_{exp} and the start state S_{st} are the same state. The bottom row indicates the consistency when we discriminate the two states.

	HSMM	HMM	ML
$(S_{\text{eva}}, \text{others})$	0.811	0.807	0.734
$(S_{\text{st}}, S_{\text{exp}}, S_{\text{eva}})$	0.639	0.621	0.437

browsing state at a timing j , q_j , is simply estimated by comparing the likelihood of the observed gaze action, x_j , using the output probabilities of the HSMM based model, λ_{HSMM} . That is, $q_t = \max_{i \in S} P(x_t = o_m | q_t = i) = b_{i,m}$.

The validity of the estimated browsing states is measured by the consistency to the browsing states of the proposed choice model. That is, the browsing states identified the proposed choice model, λ_{choice} , are considered as the “ground truth” of browsing states. Table 6.1 shows the consistency of each method. The result shows that the proposed HSMM method has the highest consistency with the ground truth. Example sequences of estimated browsing states are shown in Figure 6.10. We can see that the HSMM method replicates the ground truth of browsing states better than other methods; the HMM method missed the first evaluation state, and the ML method cannot discriminate the start state and the exploration state.

Dwells on Non-selected Candidates and Browsing States

We compare browsing states estimated by the HSMM, λ'_{HSMM} , and by the HSMM in Chapter 5, λ_{HSMM} , based on the performance of the detection of comparison among strong candidates in choice behavior. Figure 6.11 and Figure 6.12 show the proportion of dwells on the selected item, non-selected candidates, and non-candidates in each browsing state estimated by λ'_{HSMM} and λ_{HSMM} . Recall that the states s_3 and s_2 were interpreted as the representative orientation/verification state and evaluation state respectively in Chapter 5. With both tasks, the browsing states estimated by λ'_{HSMM} can separate comparison behavior of strong candidates better than ones estimated by λ_{HSMM} .

Using λ'_{HSMM} , there were significant differences of the ratio of the dwells on the selected items between S_{exp} and S_{eva} (Welch’s t-test; $p < .001$ for two-candidate-tasks and $p = 0.006$ for three-candidate-tasks). Meanwhile, we did not observe any significant difference of the ratio of the dwells on selected items

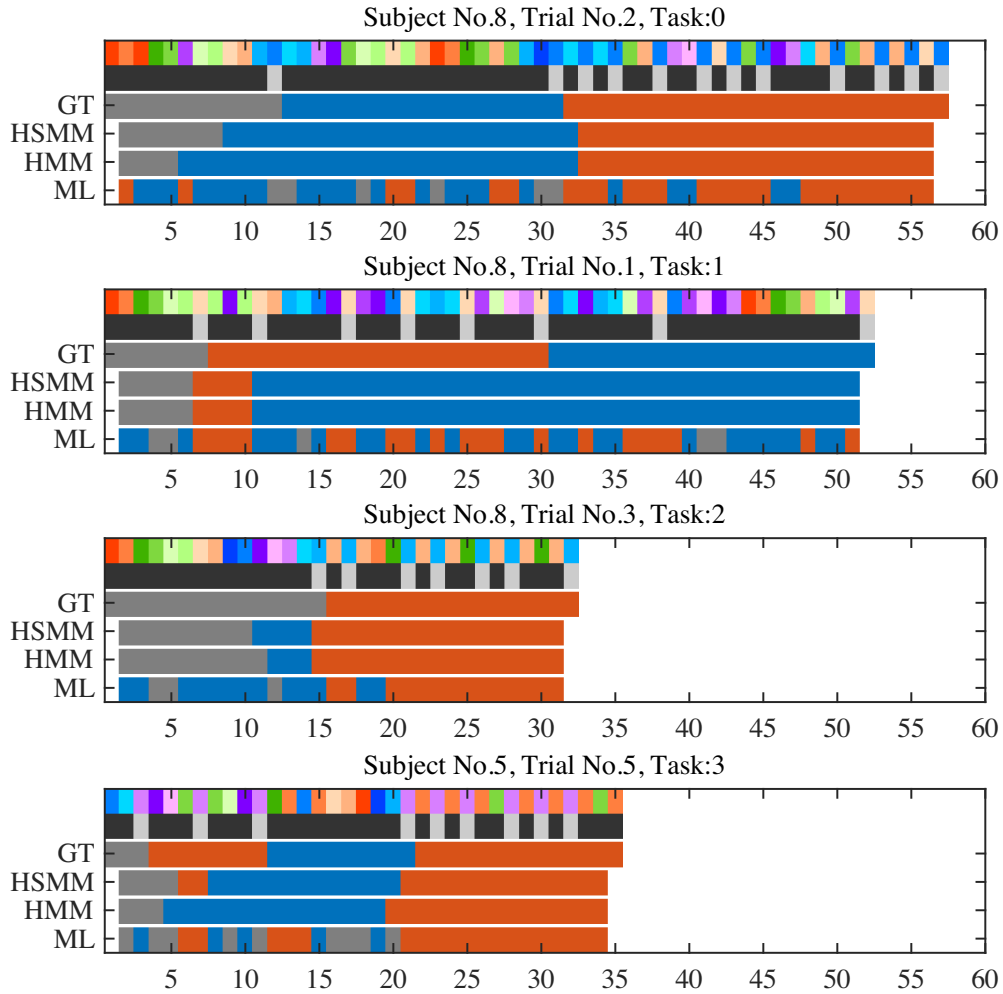


Figure 6.10: Example sequences of estimated browsing states. The blue and red intervals correspond to the exploration state, S_{exp} , and the evaluation state, S_{eva} , respectively. The gray intervals correspond to the start state, S_{st} .

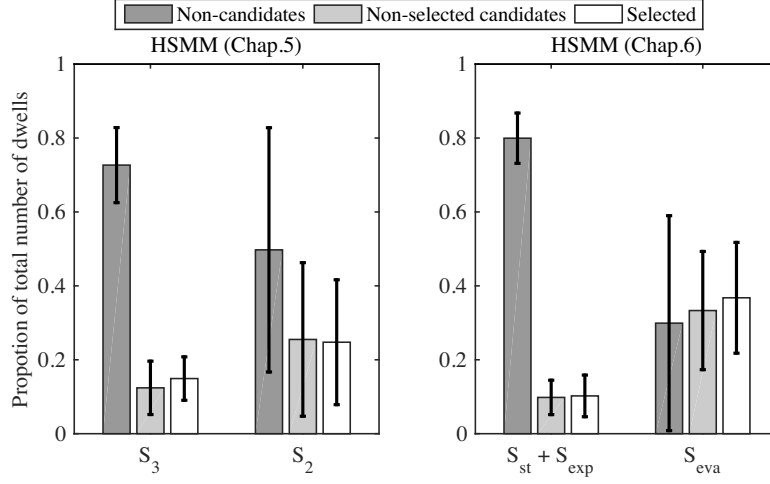


Figure 6.11: The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 2). The whiskers represent standard deviation.

between s_2 and s_3 ($p = 0.076$ for two-candidate-tasks and $p = 0.319$ for three-candidate-tasks). For dwells on non-selected candidates, there were significant differences between S_{exp} and S_{eva} ($p < .001$ for two-candidate-tasks and $p < .001$ for three-candidate-tasks) and between s_2 and s_3 ($p = .045$ for two-candidate-tasks and $p < .001$ for three-candidate-tasks). The results show that the adapted choice model, λ'_{HSM} , can identify evaluation states where strong candidates are compared and evaluated better than the unsupervised model, λ_{HSM} .

6.5 Discussion

We collected gaze data with controlled task complexities and observed differing cases of decision process in terms of browsing states. For example, some decision makers change browsing states frequently, while others just shifted from the start/exploration state to the evaluation state. This example illustrates that we cannot easily separate the decision process into two or three nicely partitioned stages. Our findings suggest that the probabilistic approach of modeling gaze behavior in choice behavior is effective since the model can adapt to the unknown complexities of the decision stages.

In this study, we investigated the relation between complexity of decision tasks (the number of strong candidates) and transitions of browsing states of a decision maker. The results of this study indicate that the evaluation state be-

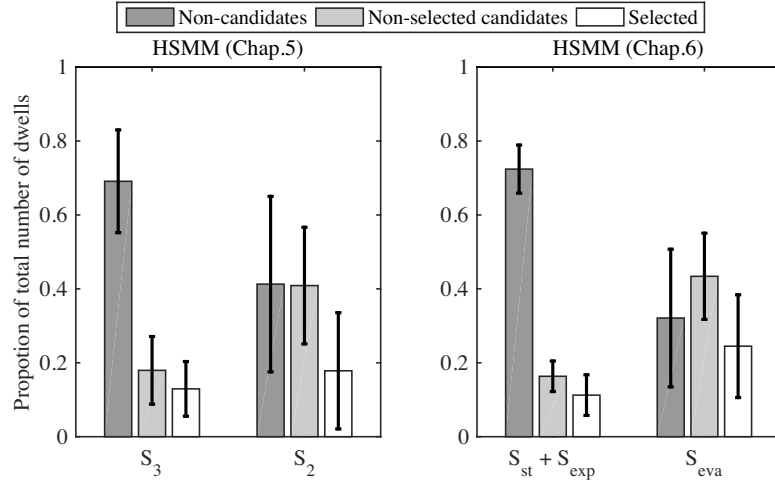


Figure 6.12: The proportion of dwells on selected items, candidates, and others in each browsing state (Number of candidates is 3). The whiskers represent standard deviation.

comes longer as task complexity increases. In this paper, we only built a single model using all collected gaze data; however, the difference between tasks we observed might indicate the necessity of the multiple models that correspond to different decision tasks. Moreover, it is also possible that the personality of decision makers and the decision strategy affect the transitions of browsing states. To improve the choice model, we need to collect and analyze more gaze data in different decision situations.

Possible Extensions of the Proposed Choice Model

The proposed choice behavior model in this paper consists of the minimum number of hidden states that can represent our assumptions: two browsing states (exploration and evaluation) and a start state. This is because we kept our model as simple as possible in order to avoid overfitting since the amount of data was limited. However, we yet do not know how many states are most appropriate to represent browsing states in choice behavior. For example, we might need to discriminate the first exploration state and interruptions of exploration states that appear later in a session. In that case, we can consider to optimize the number of states in a data-driven manner by maximizing the likelihood of the model.

Moreover, gaze behavior of decision makers is expected to be affected by exogenous factors at the beginning of decision process. For example, in our situation where a decision maker is browsing a digital catalog, the decision maker

tends to confirm all alternatives before proceeding to their detailed evaluation. The choice behavior model might have to be extended so that it can manage the effects of the exogenous factor such as the number of alternatives or layouts of catalogs.

Other Eye Movement Features to Identify Browsing States

We considered interval lengths between dwells on a selected item to identify browsing states. On the other hand, some studies of decision stage analysis have shown that there are some other cues derived from eye movements that have correlation with browsing states, e.g., fixation duration [Glöckner and Herbold, 2011]. Those cues might be useful as well for the identification. To find appropriate eye movement features for decision stage analysis, it might be possible to apply a statistical approach of feature selection using a larger gaze dataset.

Chapter 7

Conclusion

7.1 Summary

This thesis presented a novel framework to interpret the meaning of eye movements in content browsing situations. The main contribution of this study is two-fold: (1) a framework to encode eye movements based on designed structure of visual content and (2) a hierarchical content browsing model to manage the diversity of eye movement patterns. Chapter 4 presented the first contribution using designed structure defined in Chapter 3, and Chapter 2, 5 and 6 presented the second contribution using encoded gaze actions in Chapter 4.

In Chapter 3, the definition and representation of designed structure was presented. This study modeled two levels of designed structure: spatial structure of content elements (appearance-level designed structure) and their semantic relationships such as “being in the same group” (intention-level designed structure). Using the designed structure, in Chapter 4, the framework of encoding eye movements was proposed and evaluated with an experiment involving 12 participants, wherein the participants’ states were estimated from their gaze behavior. The results from the experiment show that the use of designed structure improved the performance of browser-state estimation.

Chapter 2 proposed the hierarchical content browsing model that manages the diversity of eye movement patterns. The top layer in the model represents browsing states, which corresponds to context in content browsing. To reveal the characteristics and temporal changes of browsing states in content browsing behavior, Chapter 5 proposed a probabilistic model based on hidden semi-Markov model (HSMM). The proposed browsing state model is trained and evaluated using gaze data which were collected from eight participants in an experimentally

controlled catalog browsing situation. We confirmed the validity of the estimated states through the estimation of browsers' selected items/candidates based on the estimated states.

Finally, Chapter 6 adapted the proposed content browsing model to analyze choice behavior in digital catalog browsing. We first defined and identified the browsing states using information about how often a decision maker looks at a selected item (the item finally selected by a decision maker) as a strong indicator of the consumer decision process. We confirmed that our definition of browsing states can split decision process into meaningful browsing states. Then, we trained the browsing state model proposed in Chapter 5 using eye tracking data and corresponding identified browsing states in a supervised manner. We confirmed the validity of the proposed model by measuring consistency with the identified browsing states using selected item information.

7.2 Future Work

In this section, we summarize limitations and extensions of the proposed framework and possible applications as future work.

7.2.1 Limitation and Extension of the Proposed Framework

How to realize designed structure as visual content?

The proposed framework makes a strong assumption that the content browser understands designed structure by looking at visual content. That is, intelligibility of designed structure may influence its effectiveness in encoding eye movements. We are currently investigating how to improve the impact of designed structure on browsers' behavior in more detail. The inverse problems opens the possibility of designing visual content which can give us additional clues to understanding browsers' behavior. We expect that investigating the effects of different realizations of designed structure on browsing behavior would contribute to various fields such as automated content creation and user interface design.

How to extract design structure from visual content for gaze analysis?

As mentioned in Sec. 3.2.1, this research assumes that the designed structure is given, i.e., analysts of browsing behavior also create visual content or they can

access the information of designers' intent. However, it is not always true that analysts can obtain information about designed structure in certain situations that involve a great variety of visual content such as web browsing. We are investigating strategies to obtain designed structure from existing visual content. Since it is still a difficult problem to extract semantic information from a single image, we might use the source code of digital content. Markup languages, such as HTML and XML, already have descriptions for representing relations of elements composing the visual content; therefore, it might be more feasible to extract designers' intentions from such markup languages. We might also consider using eye tracking data to generate designed structure among objects. For example, if gaze transitions occur frequently from an object A to another object B, the two objects might have an order relation. We are currently building a framework to employ the tendency of these types of gaze transitions in a probabilistic approach, which would help to automate the extraction of designed structure from visual content.

7.2.2 Applications

In this thesis, we analyzed choice behavior in catalog browsing to represent content browsing situations. We believe the proposed framework can be applied in other content browsing situations. Here, we give possible applications that we aim to accomplish in the future.

Automated Content Design Creation

By extending the proposed model, we aim to build a generative gaze model that can simulate gaze flow with given designed structure of visual content. Such generative gaze model enables automated systems to adjust content design (e.g., layouts of objects) to guide browsers' eye movements better. That is, systems can optimize content design so that the predicted gaze behavior fit the given target gaze behavior. This could be an important contribution to many fields such as automated content creation and user interface design.

In this thesis, we investigated only two types of layouts of digital catalogs: category-based and order-based. We are investigating relations between a greater variety of design structure and gaze behavior.

An Interactive Decision-aid System

As one of possible applications of the proposed framework, we consider building an information system that can provide appropriate information to browsers by estimating browsers' states/situations. For example, it might be helpful if the system can offer more detailed information about candidates when a decision maker is in the evaluation state. Since the main purpose of these experiments in this study is to verify the effectiveness of the proposed framework, only a simple gift-selecting situation has been considered. However, to build an information system such as the one described above, we need to model viewers states during complex choice behavior including mind wandering. For modeling browser' interest, we consider using findings from existing research on user state modeling into our framework [Shimonishi et al., 2013].

Bibliography

- [Baum et al., 1970] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171.
- [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 ACM Symposium on Eye Tracking Research & Applications*, volume 1, pages 83–90.
- [Borji and Itti, 2014] Borji, A. and Itti, L. (2014). Defending yarbush: Eye movements reveal observers’ task. *Journal of Vision*, 14(3):29.
- [Brandherm et al., 2007] Brandherm, B., Prendinger, H., and Ishizuka, M. (2007). Interest estimation based on dynamic bayesian networks for visual attentive presentation agents. In *Proceedings of the ACM International Conference on Multimodal Interaction*, pages 191–198.
- [Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- [Bulling et al., 2011] Bulling, A., Ward, J. A., Gellersen, H., and Troster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753.
- [Cartella et al., 2015] Cartella, F., Lemeire, J., Dimiccoli, L., and Sahli, H. (2015). Hidden semi-markov models for predictive maintenance. *Mathematical Problems in Engineering*, 2015.
- [Chen and Pu, 2011] Chen, L. and Pu, P. (2011). Users’ eye gaze pattern in organization-based recommender interfaces. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pages 311–314.

- [Clement, 2007] Clement, J. (2007). Visual influence on in-store buying decisions: An eye-track experiment on the visual influence of packaging design. *Journal of Marketing Management*, 23(9-10):917–928.
- [Eivazi and Bednarik, 2011] Eivazi, S. and Bednarik, R. (2011). Predicting problem-solving behavior and performance levels from visual attention data. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pages 9–16.
- [Gidlöf et al., 2013] Gidlöf, K., Wallin, A., Dewhurst, R., and Holmqvist, K. (2013). Using eye tracking to trace a cognitive process : Gaze behaviour during decision making in a natural environment. *Eye Movement Research*, 6(1):1–14.
- [Glaholt and Reingold, 2011] Glaholt, M. G. and Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, 4(2):125–146.
- [Glöckner and Herbold, 2011] Glöckner, A. and Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, 24(1):71–98.
- [Goldberg et al., 2002] Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., and Wichansky, A. M. (2002). Eye tracking in web search tasks: Design implications. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, pages 51–58.
- [Greene et al., 2012] Greene, M. R., Liu, T., and Wolfe, J. M. (2012). Reconsidering yarbus: A failure to predict observers’ task from eye movement patterns. *Vision Research*, 62:1–8.
- [Hirayama et al., 2011] Hirayama, T., Sumi, Y., Kawahara, T., and Matsuyama, T. (2011). Info-concierge: Proactive multi-modal interaction through mind probing. In *Proceedings of the Asia Pacific Signal and Information Processing Association Annual Summit and Conference*.
- [Inoue et al., 2016] Inoue, H., Hirayama, T., Doman, K., Kawanishi, Y., Ide, I., Deguchi, D., and Murase, H. (2016). A classification method of cooking operations based on eye movement patterns. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, pages 205–208.

- [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 Radach, R., Hyona, J., and Deubel, H., editors, *The mind's eye: cognitive and applied aspects of eye movement research*, pages 573–605.
- [Kübler et al., 2014] Kübler, T. C., Kasneci, E., and Rosenstiel, W. (2014). Submatch: Scanpath similarity in dynamic scenes based on subsequence frequencies. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, pages 319–322.
- [Li et al., 2016] Li, B., Wang, Q., Boccanfuso, L., and Shic, F. (2016). Optimality of the distance dispersion fixation identification algorithm. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, pages 339–340.
- [Liechty et al., 2003] Liechty, J., Pieters, R., and Wedel, M. (2003). Global and local covert visual attention: Evidence from a bayesian hidden markov model. *Psychometrika*, 68(4):519–541.
- [Nakano and Ishii, 2010] Nakano, Y. and Ishii, R. (2010). Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, 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.
- [Pan et al., 2004] Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., and Newman, J. K. (2004). The determinants of web page viewing behavior : An eye-tracking study. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, number 212, pages 147–154.
- [Pasupa et al., 2009] Pasupa, K., Saunders, C. J., Szedmak, S., Klami, A., Kaski, S., and Gunn, S. R. (2009). Learning to rank images from eye movements. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 2009–2016.

- [Payne, 1976] Payne, J. W. (1976). Task complexity and contingent processing in decision making : An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16(2):366–387.
- [Peters et al., 2005] Peters, R. J., Iyer, A., Itti, L., and Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research*, 45:2397–2416.
- [Pfeiffer et al., 2014] Pfeiffer, J., Meißner, M., Prosiegel, J., and Pfeiffer, T. (2014). Classification of goal-directed search and exploratory search using mobile eye-tracking. In *Proceedings of the International Conference on Information Systems*, pages 1–14.
- [Pieters and Warlop, 1999] Pieters, R. and Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *Research in Marketing*, 16(1):1–16.
- [Reutskaja et al., 2011] Reutskaja, E., Nagel, R., Camerer, C. F., and Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, 101(2):900–926.
- [Rodrigue et al., 2015] Rodrigue, M., Son, J., Giesbrecht, B., Turk, M., and Höllerer, T. (2015). Spatio-temporal detection of divided attention in reading applications using eeg and eye tracking. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*, pages 121–125.
- [Russo and Leclerc, 1994] Russo, J. and Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Consumer Research*, 21(2):274–290.
- [Russo and Rosen, 1975] Russo, J. and Rosen, L. (1975). An eye fixation analysis of multialternative choice. *Memory & Cognition*, 3(3):267–276.
- [Salvucci and Anderson, 2001] Salvucci, D. and Anderson, J. (2001). Automated eye-movement protocol analysis. *Human-Computer Interaction*, 16(1):39–86.
- [Salvucci and Goldberg, 2000] Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, pages 71–78.

- [Santella and DeCarlo, 2004] Santella, A. and DeCarlo, D. (2004). Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the ACM Symposium on Eye Tracking Research & Applications*, pages 27–34.
- [Sattar et al., 2015] Sattar, H., Muller, S., Fritz, M., and Bulling, A. (2015). Prediction of search targets from fixations in open-world settings. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 981–990.
- [Shi et al., 2013] Shi, S. W., Wedel, M., and Pieters, F. 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.
- [Shimonishi et al., 2016] Shimonishi, K., Kawashima, H., Schaffer, E., and Matsuyama, T. (2016). Tracing temporal changes of selection criteria from gaze information. In *Companion Publication of the ACM International Conference on Intelligent User Interfaces (IUI '16 Companion)*, pages 9–12.
- [Shimonishi et al., 2013] 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*, pages 41–44.
- [Simola et al., 2008] Simola, J., Salojärvi, J., and Kojo, I. (2008). Using hidden markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9(4):237–251.
- [Stüttgen et al., 2012] Stüttgen, P., Boatwright, P., and Monroe, R. T. (2012). A satisficing choice model. *Marketing Science*, 31(6):878–899.
- [Sugano et al., 2014] Sugano, Y., Ozaki, Y., Kasai, H., and Ogaki, K. (2014). Image preference estimation with a data-driven approach : A comparative study between gaze and image features. *Eye Movement Research*, 7(3):1–9.
- [Tatler et al., 2010] Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., and Velichkovsky, B. M. (2010). Yabus, eye movements, and vision. *i-Perception*, 1(1):7–27.

- [Viterbi, 1967] Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269.
- [Wedel and Pieters, 2011] Wedel, M. and Pieters, R. (2011). A review of eye-tracking research in marketing. In Malhotra, N. K., editor, *Review of Marketing Research*, number 4, chapter 5, pages 123 – 147.
- [Wedell and Senter, 1997] Wedell, D. H. and Senter, S. M. (1997). Looking and weighting in judgment and choice. *Organizational Behavior and Human Decision Processes*, 70(1):41–64.
- [Yarbus, 1967] Yarbus, A. L. (1967). *Eye movements and vision*.
- [Yonetani et al., 2012] Yonetani, R., Kawashima, H., Hirayama, T., and Matsuyama, T. (2012). Mental focus analysis using the spatio-temporal correlation between visual saliency and eye movements. *Information Processing Society of Japan Journal*, 52(12).
- [Yu and Kobayashi, 2006] Yu, S.-Z. and Kobayashi, H. (2006). Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden markov model. *Signal Processing*, 54(5):1947–1951.

List of Publications

Journal Paper

- Erina Ishikawa, Hiroaki Kawashima, and Takashi Matsuyama: Using Designed Structure of Visual Content to Understand Content-Browsing Behavior, IEICE Transactions on Information and Systems, Vol.E98-D, No.8, pp.1526-1535, 2015. doi: 10.1587/transinf.2014EDP7422
- Erina Ishikawa Schaffer, Hiroaki Kawashima, and Takashi Matsuyama: A Probabilistic Approach for Eye-tracking Based Process Tracing in Catalog Browsing, Journal of Eye Movement Research, Vol.9, No.7:4, pp.1-14, 2016. doi: 10.16910/jemr.9.7.4

International Conference (reviewed)

- Erina Ishikawa Schaffer, Hiroaki Kawashima, and Takashi Matsuyama: Estimation of Browsing States in Consumer Decision Processes from Eye Movements, The 3rd IAPR Asian Conference on Pattern Recognition (ACPR2015), Kuala Lumpur, Malaysia, 2015. doi: 10.1109/ACPR.2015.7486543
- Erina Ishikawa: Modeling Semantic Aspects of Gaze Behavior while Catalog Browsing, Proc. of the 15th ACM on International Conference on Multimodal Interaction 2013 Doctoral Consortium (ICMI2013DC), pp. 357-360, Sydney, Australia, 2013. doi: 10.1145/2522848.2532197
- Erina Ishikawa, Ryo Yonetani, Hiroaki Kawashima, Takatsugu Hirayama, and Takashi Matsuyama: Semantic Interpretation of Eye Movements Using Designed Structures of Displayed Contents, ACM International Conference on Multimodal Interaction (ICMI2012) 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction, Santa Monica, CA, USA, 2012. doi:

10.1145/2401836.2401853

Presentation (in Japanese)

- 石川恵理奈, 米谷竜, 川嶋宏彰, 平山高嗣, 松山隆司: 提示コンテンツのデザイン構造を用いた視線運動の意味理解, 電子情報通信学会技術報告, PRMU2012-60, vol. 112, no. 225, pp.47-52, 2012.10.5.
- 石川恵理奈, 川嶋宏彰, 松山隆司: 視線運動からの商品カタログ閲覧状態の推定, 第 18 回情報論的学習理論ワークショップ (IBIS2015), 2015.

Chapter 4 ©2014 IEICE. Chapter 4 is based on the work published in *Erina Ishikawa, Hiroaki Kawashima, and Takashi Matsuyama: Using Designed Structure of Visual Content to Understand Content-Browsing Behavior, IEICE Transactions on Information and Systems, Vol.E98-D, No.8, pp.1526-1535, 2015. doi: 10.1587/transinf.2014EDP7422*

Chapter 5 ©2015 IEEE. Chapter 5 is based on the work published in *Erina Ishikawa Schaffer, Hiroaki Kawashima, and Takashi Matsuyama: Estimation of Browsing States in Consumer Decision Processes from Eye Movements, The 3rd IAPR Asian Conference on Pattern Recognition (ACPR2015), Kuala Lumpur, Malaysia, 2015. doi: 10.1109/ACPR.2015.7486543*
URL: <http://ieeexplore.ieee.org/document/7486543/>