Applying Information Foraging Theory to Understand User Interaction with Content-based Image Retrieval

Haiming Liu
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA, UK
h.liu@open.ac.uk

Paul Mulholland
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA, UK
p.mulholland@open.ac.uk

Dawei Song
School of Computing
The Robert Gordon University
Aberdeen, AB25 1HG, UK
d.song@rgu.ac.uk

Victoria Uren
Department of Computer
Science
University of Sheffield
Milton Keynes, MK7 6AA, UK
v.uren@dcs.shef.ac.uk

Stefan Rüger
Knowledge Media Institute
The Open University
Milton Keynes, MK7 6AA, UK
s.rueger@open.ac.uk

ABSTRACT

The paper proposes an ISE (Information goal, Search strategy, Evaluation threshold) user classification model based on Information Foraging Theory for understanding user interaction with content-based image retrieval (CBIR). The proposed model is verified by a multiple linear regression analysis based on 50 users' interaction features collected from a task-based user study of interactive CBIR systems. To our best knowledge, this is the first principled user classification model in CBIR verified by a formal and systematic qualitative analysis of extensive user interaction data.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.1.2 [**User/Machine Systems**]: Human factors

General Terms

Human Factors, Theory, Verification

Keywords

Information Foraging Theory, User Interaction, Exploratory Search, User Modelling, Content-based Image Retrieval

1. INTRODUCTION

Learning more about users is vital to improve their interaction with CBIR systems especially through relevance feedback (RF) [6]. User interaction involves three key elements: a user interaction model, an interactive interface for delivering the user interaction model, and users. The three elements combine to enable effective interaction to happen.

To make relevance feedback mechanisms more interactive, some researchers have focused on developing user interaction models to formalize different factors for improving the interaction. For example, Spink et al. (1998) [15] proposed a three-dimensional spatial model to support user interactive search for text retrieval. Campbell (2000) [1] focused on the time dimension, and proposed the Ostensive Model (OM) that indicates the degree of relevance relative to when a user selected evidence from the results set. Ruthven et al. (2003) adapted two dimensions from Spink et al.'s model and combined with OM in their study [14]. Liu et al. (2009) [7] introduced a four-factor user interaction model (FFUIM) for content-based image retrieval (CBIR), which includes relevance region (relevant and non-relevant), relevance level (how relevant or non-relevant), time and frequency (of an image being selected as a relevant/non-relevant example).

The user interaction models aim to improve the interaction between users and the systems and in turn users' overall search experience. The models need to be delivered by visual interactive interfaces for supporting users in grasping how RF algorithms work and how they can be manipulated.

Some visual interactive interfaces have been developed to deliver the models and to further improve the user interaction. For instance, Urban et al. (2006) developed an image search system based on the OM [18]. Liu et al. (2009) proposed an interactive CBIR interface that successfully delivered the FFUIM and allowed users to manipulate the model effectively [8].

We consider users' interaction through the interactive interfaces as a searching, learning and investigating process, which is exploratory in nature. Exploratory search, emerging as a new paradigm of information seeking, aims to shift the research focus from getting the highest precision (query-document matching) toward finding guidance at all stages of the information seeking process to support a broader set of users' searching and interaction behaviors [9, 19].

White and Roth (2009) [19] suggested exploratory search is related to Information Foraging Theory (IFT) [12, 13] in the respect of finding an optimal patch to reach the users' information goal during search, depending on various user-oriented factors, such as how clear users' information goal

is, how users apply their searching strategy, and how users decide which information to use, etc.

Information Foraging Theory (IFT) was developed by Pirolli et al. [11–13], which suggests the way humans seek information is like the way wild animals gathering food. Animals' general food seeking behavior is that they will first find a path of food resource (scents); next they will select what to eat (diet); and then they have to decide when to hunt elsewhere (patch) [16].

Mulholland et al. (2008) [10] showed that IFT can interpret the effects of the exploratory search technologies. They identified two distinct user strategies of exploratory search, namely, risky search strategy and cautious search strategy. Their findings are a concrete step forward in supporting exploratory search.

The users themselves are also a key element of the interaction. Users can be very different when they use search systems: some people know what they want, and some only know when they find it [17]; some people are patient, but some are not; some people frequently change their mind on what they are looking for, but some do not [18].

In this paper, we aim to investigate a model for categorizing different user types based on IFT. The model will be verified qualitatively by a systematic analysis of the extensive user interaction data collected from a real interactive image search scenario.

2. INFORMATION FORAGING THEORY

To provide adaptive strategies for information foraging in a complex information environment, Pirolli and Card (1995) [12] proposed IFT, which aims "*to explain and predict how people will best shape themselves for their information environments and how information environments can best be shaped for people.*" *(P.3)* [11]. The methodology of IFT is adapted from the framework of optimal foraging theory in biology [16]. Pirolli and Card [13] adapted two conventional models from optimal foraging theory [16], originally applied to the food hunting environment, to the information seeking environment. Further, they proposed three information models for IFT: information patch model, information diet model and information scent model.

2.1 The Information Patch Model

The aim of the information patch model is to predict the amount of time a forager would spend within an information patch before searching for new patches. This is important when information is distributed in number of patches. For instance, there are a variety of information items in my office, such as books, printed papers, notes, electronic files in my computer and an internet connection. Some of these items are located within arms reach of my desk, and some items are stored on the bookcase or in the filing cabinet. The relevant information to my current task can be found on the desk and on the bookcase. If I identify the arms reachable area as one patch and the bookcase as another patch, my information foraging process will comprise within-patch and between-patch activities. I will need to decide whether I search longer on the desk or I should go to search the relevant information on the bookcase. The decision will be made depending on the prevalence and profitability of the patches. A higher prevalence of patches may contain many relevant items to the task, and a more profitable patch may contain the most relevant information to the task. All in all,

the decision on when to do within-patch and between-patch activity will depend on the user's judgement as to which approach will complete the task in the shortest time.

2.2 The Information Diet Model

The question that the conventional diet model deals with is: when a predator lives in an environment containing a number of potential kinds of food sources, what kinds of things should the predators prey on, and what kinds of things should they ignore? One way to answer this question is in terms of the diet concept: a generalized diet includes a broad type of prey, but a specialized diet includes only a few types. "*If a predator is too specialized, it will do very narrow searching. If the predator is too generalized, then it will pursue too much unprofitable prey (p.39)* [11]." Thus, the diet model in IFT can be explained in terms of the conventional diet model: if I have a generalized diet, I will complete the task with a wide range of relevant information with diverse dimensions; if I have a specialized diet, I will complete the task with only a few relevant information sources having precise characteristics.

2.3 The Information Scent Model

Information scent model is a psychological theory, which explains how people identify the value of information based on cues, such as result clusters shown on the interface in order to gain an overall sense of the information space. If the scent is strong, the forager will be able to move fairly directly. If there is no scent, the forager will perform a random walk [13].

The information scent model has been applied to investigate effective information scent cues in aiding navigation. For instance, Chi et al. (2001) proposed two computational methods for modeling users' information needs and actions on the Web, based on the concept of information scent. The first situation was to predict users' surfing pattern given users' information needs. The second situation was to infer users' information need given user's particular pattern of surfing. Their general finding was that the two models will help researchers better understand the usage of the Web, help in designing of better web sites, and make users information seeking activities more efficient [2].

Information Foraging Theory (IFT) has been suggested and applied to improve the design perspective for interactive search and in turn to improve users search experience [4, 5, 9, 11, 19]. However, there is a lack of research on applying IFT to understand user interaction from users' perspective.

Based on the experience from our task-based users study and motivated by the findings in Mulholland et al. (2008) [10], we consider that users can be classified into different user types based on their interaction profiles of search preferences and search behaviors. We then decide to investigate in depth how many different user types we can get and what search preferences and behaviors each user type has, based on IFT and a qualitative analysis on the real user interaction data for interactive search.

3. DEFINITION OF ISE MODEL BASED ON INFORMATION FORAGING THEORY

In this section, we propose a new user classification model, called ISE model, based on IFT. The model includes three criteria: information goals (I), search strategies (S) and evaluation thresholds (E). Each criterion categorizes users into

two different user characters[1]: I - fixed information goal or evolving information goal; S - risky search strategy or cautious search strategy; E - weak evaluation threshold or precise evaluation threshold. We take a task-based image interactive search scenario (Section 4.2) as an example to explain how we map between the three information models of IFT and the three criteria of the ISE model. The information patch in this scenario will be a set of result images. The information scent will be the clues that users learn from task descriptions, query images and result images to formulate their information goal and manage their search process. The information diet will be the way users select the feedback and result images.

3.1 Information Goal (I)

The information goal can be explained by the information scent model of IFT.

After reading the task description, the searchers might or might not have a clear information goal (idea on what they are looking for) to start the search. In IFT terms, the searchers might or might not get strong information scent from reading the task based on their information environment (knowledge). Thus, the searchers can be categorized into two types based on the information scent model: one type with fixed information goal and the other with evolving information goal. According to the information scent concepts, if the searchers have a fixed information goal, they will focus on what they are looking for and will likely to make consistent decisions at every stage. On the other hand, if the searchers have an evolving information goal, their search will be more exploratory. They will randomly walk through and learn from the data before they make a decision.

The hypothesis with respect to the information goal for interactive image search are: (1) the searchers with evolving goal will be likely to do trial-and-error types of search so that it will take them longer to find the best results for completing search tasks. For example, they will reformulate queries with completely different image examples, and they may go back to previous queries if they change their information goal; (2) the searchers with fixed goal are likely to have an opposite behavior to the searchers with evolving goal. For instance, they will refine queries with small changes to the image examples in the queries, and they are likely to get consistent results with every query refinement, so that they do not need to reuse previous queries and are likely to get satisfying result images quickly for completing the search tasks.

3.2 Search Strategy (S)

The search strategy can be explained by the information patch model of IFT.

When the searchers start the search, they will submit the first query, which can be seen as an initial effort to find the first information patch, and then they might or might not walk around within the patch and evaluate what they have found before they provide feedback to refine or reformulate the query and start a new search (we can consider this as looking for a new patch). In IFT terms, the searchers can decide whether they would like to perform between or within patch activities based on their search strategy. Thus, we can categorize the searchers into two types based on the

---

[1]There are in total six characters in the ISE model.

information patch model. Motivated by the findings of Mulholland et al. (2008) [10], we suggest that the one type of searchers will have cautious search strategy and the other have a risky search strategy. According to the information patch concept, the searchers with cautious search strategy will do more within-patch activities, which means they will carefully search through the current patch before they go to the next patch (e.g. refining the query to start a new search); the searchers with risky search strategy, on the other hand, will be more adventurous and perform more between-patch activities, meaning that they will skip over the current patch and move to next patch quickly.

The hypotheses underlying the search strategy for interactive image search are: (1) the searchers with cautious search strategy will look through the search result very carefully page by page and spend a long time to analyze the results before they refine the query to start a new search, and they will not select the result images until they think no better images exist in the result set; (2) the searchers with risky strategy will only look at the first few pages and select the result images from the pages while they are viewing, and then they will reformulate a new query to start another search.

3.3 Evaluation Threshold (E)

The evaluation threshold can be explained by the information diet model of IFT.

When searchers select the result images for completing the tasks, they need to decide which images to choose. In IFT terms, some foragers like easy-to-catch prey, but others like hard-to-catch prey. Thus, the searchers can be categorized into two types based on the information diet concepts: one type with weak evaluation threshold and the other with precise evaluation threshold. According to the information diet concepts, the searchers with weak evaluation threshold will be likely to go for easy-to-catch information although the information maybe just slightly relevant to the their information goal; the searchers with precise evaluation threshold will instead go for hard-to-catch information: for example, they will not select the information unless it is highly relevant to their information goal.

The hypotheses of evaluation threshold for interactive image search are: (1) the searchers with weak evaluation threshold will select a large number of images based on diverse relevance to their search information goal, for example, if they are looking for a picture of apple, they will be happy with any picture as long as there is an apple on the picture; (2) the searchers with precise evaluation threshold will only select highly relevant images to their search information goal, for instance, if they are looking for a picture of apple, they will not select an image unless there is a red apple in the image, and they will refine the query carefully and try to achieve the precise result.

In summary, Table 1 shows the mapping between IFT and ISE model (including three criteria and six characters), and Table 2 shows the definitions of the six characters in ISE model.

4. VERIFICATION OF ISE MODEL BY QUALITATIVE DATA ANALYSIS

The above definitions of the six characters in ISE model are based on the mapping between IFT and the interactive image search scenario. In order to verify ISE model, we first

Information Foraging Theory	Criterion	Character
Information scent models	information goal	fixed evolving
Information patch models	search strategy	cautious risky
Information diet models	evaluation threshold	weak precise

Table 1: ISE user classification model based on the Information Foraging Theory

Character	Definition
fixed	Searchers with fixed information goal know what they are looking for.
evolving	Searchers with evolving information goal are not sure what they are looking for.
cautious	Searchers with cautious search strategy move slowly between patches.
risky	Searchers with risky search strategy move quickly between patches.
weak	Searchers with weak evaluation threshold are lenient on selecting the result.
precise	Searchers with precise evaluation threshold are strict on selecting the result.

Table 2: Definitions of the six user characters

operationalise the definitions of the six user characters by mapping them to concrete user interaction features collected from the screen capture of an extensive user study that we have performed, and then verify the model by a formal and systematic qualitative analysis of the interaction features.

4.1 The Verification Procedure

The procedure of ISE model verification is as below:

1. extract the user interaction features from the screen captures of a task-based user study;

2. operationalise the definitions of the six characters in ISE model based on the interaction features;

3. apply a multiple linear regression test to the interaction features;

4. check whether the regression models reasonably enable the operational definitions of the six characters in ISE model.

4.2 User Study

Our user study from which the user interaction data used for the analysis in this paper were collected, was to investigate the effects of a novel user interaction framework called 'uInteract' for interactive CBIR. The uInteract framework includes a FFUIM [7] and an interactive interface [8]. In the study, we employed a total of 50 users for three focused evaluations of 12 interactive CBIR systems (various interaction models and the supporting interfaces). The users were a mixture of males and females, undergraduate and postgraduate students and academic staff from a variety of departments. The 50 users were divided into three groups, and 17, 16 and 17 users were assigned to the three focused evaluations respectively. In each evaluation, the users were given four search tasks with different complexity levels to complete in a time constraint, on four interactive CBIR systems in a random order, and provided feedback on their search experiences through questionnaires and informal interviews. The whole process was recorded by screen capture

with video and audio input. The ImageCLEFphoto2007 collection [3] was used, which consists of 20,000 real life images and 60 query topics. The interfaces of the systems provided varying functionalities to support users' exploratory search, such as, query history, negative query, showing negative result, query image importance scoring, etc. The evaluations applied simulated natural life tasks, such as *"Imagine you are a graphic designer with responsibility for the design of leaflet on the newly built sport stadium for the local council. The leaflet is intended to raise interest among the general public and encourage people to use the stadium and to watch the sports in the stadium. Your task is to find 3 5 images, from a large collection of images, to include in the leaflet. The images should represent the kind of sports you think can be held in the stadium.",* which allowed the users to develop their own interpretation of the task description, use their own judgement for choosing relevant images as feedback and result, and decide when to use different functionalities on the interface to support their search.

The goal of evaluation one (E1) was to test whether users find the uInteract interface (Figure 1) [8] useful and easy to use. The goal of evaluation two (E2) and three (E3) is to test the performance of the four profiles of the OM [1], and the four settings of the FFUIM [7], respectively, under the uInteract interface.

4.3 The User Interaction Data

A substantial amount of real user interaction features are extracted from the screen capture of the three evaluations in our user study. There are in total 50 users' screen captures. Every screen capture is about two hours long with both audio and video inputs. We extract totally 123 interaction features from the screen captures (37 from E1, 44 from E2, 42 from E3). Figure 2 shows the 123 interaction features and their descriptive mean values in the three evaluations. Some interaction features apply to more than one evaluation. There are 48 unique interaction features within the total 123 interaction features. Figure 3 shows the 48 unique interaction features and their descriptions.

We can roughly categorize the 48 unique interaction features into six groups:

- **time and iteration:** time to complete each iteration, time to complete task, time to find the best result, number of iterations/queries per task;

- **results page:** number of result pages viewed, page result selected from, page found the best result[2], page positive feedback selected from, page negative feedback selected from;

- **image:** number of images per query (positive and negative query), number of feedback images selected (positive and negative query), number of results selected;

- **functionality used:** number of times positive/negative ranking used, number of times positive/negative history used;

- **select result strategy:** some users select result while searching, and some users select result at the end of the search;

---

[2]The best result here is judged based on the rating result of the five independent raters.
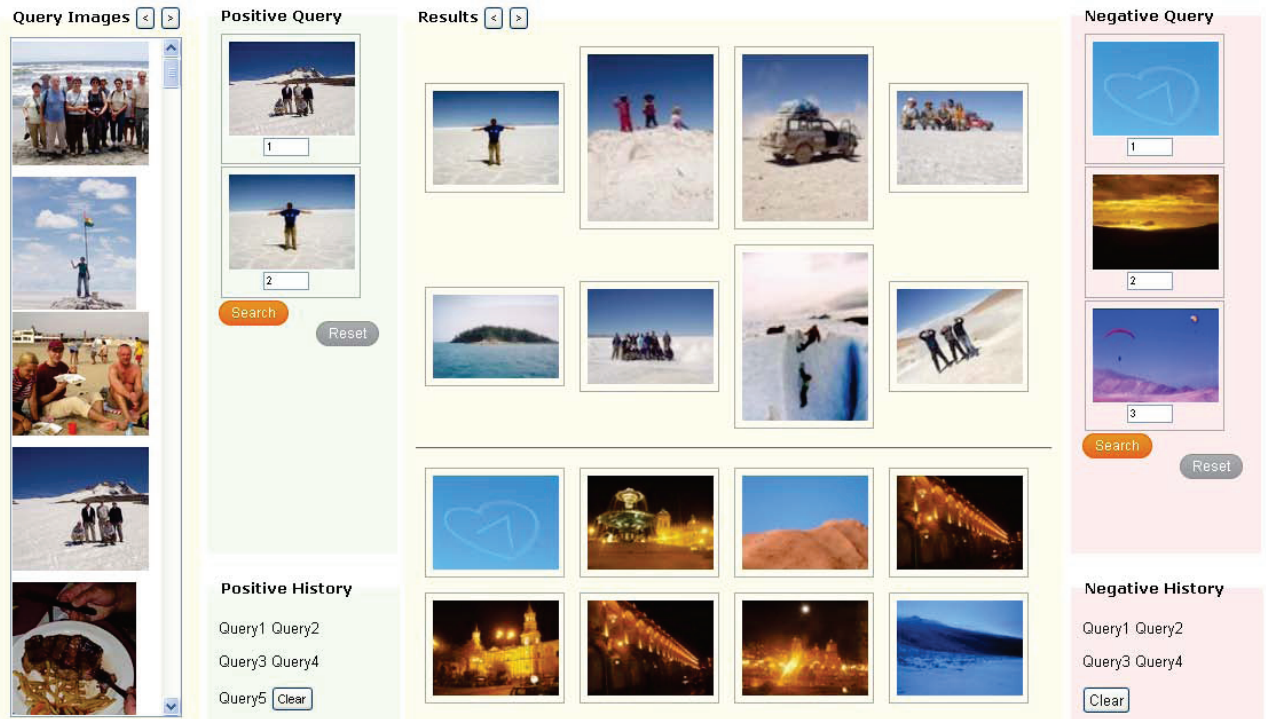
Figure 1: The uInteract interface

Query transition	description
Repeat	consecutive positive or negative query contains identical images.
Subset	The next positive or negative query contains a subset of the query images.
Superset	The next positive or negative query contains all the previous images plus one or more additional images.
Overlap	The next positive or negative query contains some but not all of the previous images plus one or more additional images.
Jump	There is no intersection between the images used in consecutive positive or negative queries.

Table 3: Five kinds of query transition

Characters	Operational definition
Fixed	1. use small number of jump query transition; 2. use small number of history functionality; 3. find the best result image early.
Evolving	1. use large number of jump query transition; 2. use large number of history functionality; 3. find the best result image late.
Cautious	1. view large number of result pages; 2. spend long time per search iteration; 3. select result at the end of the search.
Risky	1. view small number of result pages; 2. spend short time per search iteration; 3. select result while searching.
Weak	1. select large number of results; 2. select large number of feedbacks; 3. search small number of iterations.
Precise	1. Use lots of subset query transition; 2. use many times ranking functionality; 3. search large number of iterations.

Table 4: Operational definitions of the six characters

- **query transitions:** we adapted the five query transitions from Mulholland et al (2008) study [10] to our analysis. The five transitions for both positive and negative queries are in Table 3.

4.4 Operationalizing ISE Model Based on Interaction Features

After defining ISE model (including three criteria and six user characters) based on IFT, we examine the 48 unique interaction features from the collected user interaction data, and assign the six characters or their combinations to every interaction feature based on the definitions provided in Table 2. Figure 4 gives an example of how and why the assignments were made.

Table 4 summarizes the operational definitions of the six characters that we proposed based on the character allocations to 48 unique interaction features. We can then verify the ISE model by qualitative data analysis on the 123 interaction features based on the operational definitions.

4.5 Multiple Linear Regression

Multiple linear regression is applied to the qualitative data analysis, because we want to find out the correlations across the interaction features, and also want to generate optimal regression models for predicting the interaction features. The 123 interaction features are the input of the multiple linear regression. We carry out the regression test by SPSS, a statistical analysis tool.

We first apply the multiple linear regression on the interaction features of the three evaluations respectively. We then get a model to predict each interaction feature. For example, in Figure 5, interaction feature $TimePerIteration$
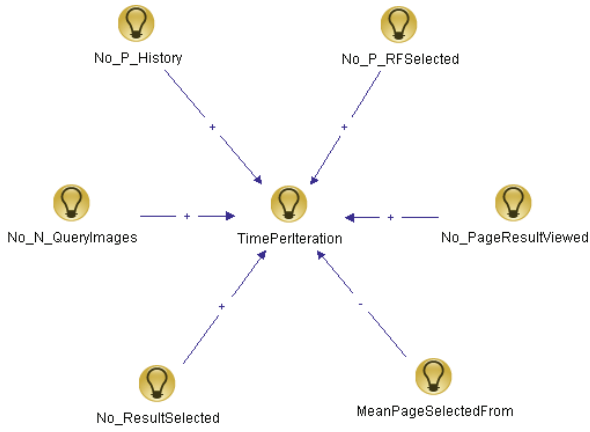
Figure 5: An example of visualized multiple linear regression model

Character	No. of models
Cautious	7
Risky	2
Evolving	7
Fixed	1
Weak	12
Precise	12
Cautious+Evolving	2
Cautious+Weak	2
N/A	78
Total	123

Table 5: Summary of characters and no. of supporting regression models

is predicted by $No\_P\_RFselected$, $No\_PageResultViewed$, $No\_N\_QueryImages$, $No\_ResultSelected$, $No\_P\_History$ and $MeanPageSelectedResultFrom$ interaction features. There are relation lines between every interaction feature and the predicted feature. The direction of each arrow is the predicting direction. The $+$ or $-$ on the line means the prediction is positive or negative. For instance, the $No\_PageResultViewed$ predicts $TimePerIteration$ positively. In other words, if a user view more result pages, they are likely to spend longer time per search iteration.

4.6 Regression Model Analysis

We have obtained the 123 regression models for the 123 interaction features from the three evaluations. We then need to investigate whether the regression model are consistent with the operational definitions of the proposed six user characters.

We assign the six user characters in the ISE model and their combinations to the 123 regression models. The justification method for assigning a character to a model confidently is that the model has to contain at least two interaction features that are relative to the character's operational definition of ISE model (Table 4). Examples of the assignments of characters to regression models are shown in Figure 6[3].

The results show that the models is consistent with the the six characters or their combinations, and the descriptions fit[4] the operational definitions of the six characters. Take the regression model predicting $No\_PageResultViewed$ in Figure 6 as an example. $No\_PageResultViewed$ is positively predicted by $TimePerIteration$, $No\_ResultSelected$, $MeanPage$-$SelectedFrom$ and $No\_N\_QueryImages$, which means that users will view lots of result pages if they spend a long time per search iteration, select a large number of result images, select result images from late pages, and use large num-

ber of negative query images. According to the operational definitions and our justification method, this model can be described by cautious character because the model contains two interaction features ($No\_PageResultViewed$ and $TimePerIteration$) that are related to the operational definition of the cautious character, and the description of the regression model fits the operational definition well, e.g., spends long time per search iteration ($TimePerIteration$) and views large number of result pages ($No\_PageResultViewed$).

From the 123 regression model analysis results we can see that all the predictions in the models are reasonable and most of the regression models are consistent with the corresponding characters' definitions. Some regression models can be described by single character but some models need to be described by different combinations of the six characters. Some characters correspond to a large number of regression models but some only correspond to a couple of regression models. The user characters and their combinations corresponded to at least one regression models are listed in Table 5[5].

Table 5 shows that the 45 regression models confidently correspond to 8 character groups based on our justification method[6]. Each character group[7] is identified by on average 5.6 regression models. We consider that the four character groups identified by more than six regression models are well represented character groups in our user study. They are *cautious*, *evolving*, *weak* and *precise*. Seventy eight regression models can not be covered by any character groups based on our justification method. Within the 78 regression models, 8 models do not include any interaction features that are relative to the operational definitions of the six characters in ISE model, but 70 models contain one/more single interaction feature that is/are relative to the operational definition of one/more characters. According to the judgement method for ISE model verification, the 70 regression models should not be used to verify the ISE model, although the models are all consistent with the operational

---

[3]The "$+/-$" shows how the features in the regression models predict the interaction features in the second column. "$+$" means the prediction is positive, and "$-$" means the prediction is negative. "$*$" indicates that the interaction features are not mentioned in the operational definitions in Table 4.
[4]"Fit" means all the mentioned features in a regression model have correct detection based on the operational definitions.

---

[5]In Table 5, $N/A =$ there is no more than one interaction feature relative to the operational definition of any character in the regression model.
[6]The interaction features in the regression models show reasonable predictions. Further, there are at lease two interaction features in each model that fit the operational definitions of the six characters in ISE model.
[7]The character group can be any single character or the characters' combinations. We consider a character group is a user type.

definition. Therefore, we only take into consideration the 45 models that contain at least two interaction features that are relative to the characters' operational definition.

5. CONCLUSIONS

In an effort to understand the users' interaction in regard to different user types for CBIR, we have proposed a user classification model - ISE - based on Information Foraging Theory. The ISE model contains three criteria: information goal (I), search strategy (S) and evaluation threshold (E). There are different types of user characters in each criterion. They are fixed information goal and evolving information goal (I); risky search strategy and cautious search strategy (S); weak evaluation threshold and precise evaluation threshold (E).

In order to verify the ISE model, we have operationalized the ISE model based on the 48 unique interaction features gathered from the screen capture of in a user study on various interactive CBIR systems. A multiple linear regression has then been performed on the total number of 123 interaction features from the three evaluations in the user study, resulting in 123 regression models. Finally, we have investigated whether the regression models are consistent with the operational definitions of the six user characters in ISE based on a thorough analysis of the regression models.

The ISE user classification model has been successfully verified by the qualitative data analysis. The findings show that all regression models are consistent with the operational definitions of the six characters in the ISE model. Eight user character groups (user types) are confidently identified by 45 regression models.

This practice has not only helped to find different user types for future user-focused design, study and analysis, but also reinforced the usefulness of IFT for interactive CBIR search, and for exploratory search in general.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] I. Campbell. Interactive evaluation of the ostensive model using a new test collection of images with multiple relevance assessments. *Journal of Information Retrieval*, 2(1), 2000.

[2] E. H. Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions and the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 490–497, New York, NY, USA, 2001. ACM.

[3] M. Grubinger, P. Clough, H. Müller, and T. Deselaers. The iapr tc-12 benchmark: A new evaluation resource for visual information systems. In *In Proceedings of International Workshop OntoImageŠ2006 Language Resources for Content-Based Image Retrieval*, pages 13–23, 2006.

[4] M. Käki. *Enhancing Web Search Result Access with Automatic Categorization*. PhD thesis, Faculty of Information Sciences, Department of Computer Sciences, University of Tampere, Finland, 2005.

[5] B. Kules and B. Shneiderman. Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing and Management*, 44(2):463–484, 2008.

[6] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Cimmunications and Applications*, 2(1):1–19, February 2006.

[7] H. Liu, V. Uren, D. Song, and S. Rüger. A four-factor user interaction model for content-based image retrieval. In *Proceeding of the 2nd international conference on the theory of information retrieval (ICTIR)*, 2009.

[8] H. Liu, S. Zagorac, V. Uren, D. Song, and S. Rüger. Enabling effective user interactions in content-based image retrieval. In *Proceedings of the Fifth Asia Information Retrieval Symposium (AIRS)*, 2009.

[9] G. Marchionini and R. W. White. Beyond search: Information seeking support systems. *IEEE Computer*, 42(3):30–32, March 2009. Guest Editor's Introduction.

[10] P. Mulholland, Z. Zdrahal, and T. Collins. Investigating the effects of exploratory semantic search on the use of a museum archive. In *Proceeding of the IEEE 2008 International Conference on Distributed Human-Machine Systems*, 2008.

[11] P. Pirolli. *Information Foraging Theory Adaptive Interaction with Information*. Oxford University Press, Inc, 2007.

[12] P. Pirolli and S. Card. Information foraging in information access environments. In *CHI '95: Mosaic of Creativity*, pages 51–58, New York, NY, USA, 1995. ACM.

[13] P. Pirolli and S. K. Card. Information foraging. *Psychological Review*, 106:643–675, 1999.

[14] I. Ruthven, M. Lalmas, and K. van Rijsbergen. Incorporating user search behaviour into relevance feedback. *Journal of the American Society for Information Science and Technology*, 54(6):528–548, 2003.

[15] A. Spink, H. Greisdorf, and J. Bateman. From highly relevant to not relevant: examining different regions of relevance. *Information Processing Management*, 34(5):599–621, 1998.

[16] D. W. Stephens and J. R. Krebs. *Foraging Theory*. Princeton University Press, 1986.

[17] A. ter Hofstede, H. Proper, and T. van der Weide. Query formulation as an information retrieval problem. *The Computer Journal*, 39(4):255–274, September 1996.

[18] J. Urban, J. M. Jose, and K. van Rijsbergen. An adaptive technique for content-based image retrieval. *Multimedia Tools and Applications*, 31:1–28, July 2006.

[19] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query - Response Paradigm*. Morgan and Claypool Publishers, 2009.

	Interaction features of Evaluation1 (Descriptive Mean)	Interaction features of Evaluation2 (Descriptive Mean)	Interaction features of Evaluation3 (Descriptive Mean)
1	No_N_QueryImages (4.56)	No_P_QueryImages (3.82)	No_N_RFSelected (3.20)
2	MeanPageSelectedResultFrom (2.78)	MeanPageResultSelectedFrom (5.65)	MeanPageResultSelectedFrom (4.30)
3	TimePerIteration (0:52)	MeanPage_P_RFSelected (3.46)	TimePerIteration (1:23)
4	No_ResultSelected (0.5)	No_ResultSelected (2.36)	No_PageResultViewed (6.77)
5	No_PageResultViewed (5.09)	No_N_QueryImages (3.27)	No_P_RFSelected (2.26)
6	MeanPage_P_RFselected (3.20)	TimePerIteration (1:40)	MeanPage_P_RFSelected (3.18)
7	SelectResultsWhileSearching (0.32)	No_PageResultViewed (7.29)	No_N_QueryImages (4.26)
8	No_P_SubsetQuery (1.19)	MeanPage_N_RFselected (3.54)	PageN_RFSelected (3.78)
9	No_P_UniqueimagesPerTask (8.47)	TimeFindBestResultImage (3:48)	No_ResultSelected (2.33)
10	No_N_UniqueimagesPerTask (5.34)	No_P_RepeatQuery (1.77)	MinTimePerIteration (1:09)
11	PageFindBestResultImage (1.22)	No_N_RepeatQuery (2)	No_P_Q_ImagesPerTask (18.85)
12	No_P_RepeatQuery (1.05)	No_N_SupersetQuery (0.9)	No_P_RepeatQuery (1.41)
13	TotalPageNRFSelected (2.13)	No_N_OverlapQuery (0.04)	No_P_SubsetQuery (0.37)
14	No_Iteration_Query (6.39)	No_N_JumpQuery (1.50)	No_N_RepeatQuery (2)
15	MeanNo_N_ImagesPerQuery (2.43)	SelectResultsStrategy (1.27)	No_N_UniqueimagePerTask (5.20)
16	MaxTimePerIteration (2:15)	No_N_UniqueimagesPerTask (4.08)	Mean_No_N_ImagesPerQuery (2.75)
17	TimePerIterationRange (1:45)	TotalNoResultSelected (10.91)	TotalNoNRanking (0.60)
18	TotalNoNRFSelected (1.39)	TotalPagePRFSelected (7.20)	TotalNoPageResultViewed (28.16)
19	No_P_OverlapQuery (0.5)	No_P_Q_ImagesPerTask (17.81)	SelectResultStrategy (1.31)
20	No_N_OverlapQuery (0.07)	No_P_SupersetQuery (1.13)	TotalNoPRanking (1.97)
21	TimePerTask (5:33)	No_Iteration_Query (4.66)	TotalNoPHistory (0.20)
22	TotalNoResultSelected (3.20)	TotalNoNRanking (0.30)	MeanTimePerIteration (1:59)
23	No_N_SubsetQuery (0.29)	TotalNoPHistory (0.41)	TimeFindBestResultImage (3:39)
24	TotalNoPRanking (0.98)	MaxTimePerIteration (4:19)	No_P_SupersetQuery (1.29)
25	MeanTimePerIteration (1:10)	No_P_OverlapQuery (0.38)	P_OverlapQuery (0.06)
26	MedianTimePerIteration (1:02)	TimePerTask (7:49)	No_N_Q_ImagesPerTask (12.27)
27	TotalNoPageResultViewed (32.27)	MeanNo_N_ImagesPerQuery (2)	TotalNoNHistory (0.06)
28	TotalPageSelectedFrom (9.78)	TotalNoNHistory (0.08)	TotalPagePRFSelected (9.63)
29	TimeFindBestResultImage (4:20)	TotalNoNRFSelected (3)	TotalNoNRFSelected (3.96)
30	TotalPagePRFSelected (12.63)	TotalNoPRanking (1.34)	MedianTimePerIteration (1:53)
31	No_N_RepeatQuery (2.86)	MedianTimePerIteration (1:54)	MaxTimePerIteration (3:05)
32	TotalNoPRFSelected (4.44)	MinTimePerIteration (1:13)	No_N_SubsetQuery (0.08)
33	No_N_Q_ImagePerTask (24.29)	TotalNoPageResultViewed (33.81)	No_P_UniqueimagesPerTask (6.07)
34	TotalNoNRanking (0.03)	PageFindBestResultImage (3.78)	Mean_No_P_ImagesPerQuery (4.40)
35	No_P_SupersetQuery (2.30)	MeanNo_P_ImagesPerQuery (4.40)	TotalNoResultSelected (9.85)
36	No_P_Q_ImagesPerTask (29.36)	TotalPageSelectedFrom (48.02)	TotalPageSelectedFrom (42.04)
37	Mean_No_P_ImagesPerQuery (4.74)	No_N_Q_ImagesPerTask (10.92)	TotalNoPRFSelected (3.19)
38		TotalPageNRFSelected (5.44)	No_N_SupersetQuery (0.63)
39		MeanTimePerIteration (2:15)	TimePerTask (5:48)
40		No_P_SubsetQuery (0.34)	TimePerIterationRange (1:56)
41		No_P_JumpQuery (0.13)	No_N_JumpQuery (1.06)
42		No_N_SubsetQuery (0.23)	No_Iteration_Query (4.16)
43		No_P_UniqueimagesPerTask (5.48)	
44		TotalNoPRFSelected (2.58)	

Figure 2: 123 interaction features generated for Evaluations 1, 2 and 3 and the features' descriptive means

	Interaction feature	Description
1	No_N_QueryImages	Number of images in a negative query per iteration
2	MeanPageSelectedResultFrom	Mean page number the result images selected from per iteration
3	TimePerIteration	Time used per search iteration
4	No_ResultSelected	Number of result images selected per iteration
5	No_PageResultViewed	Number of result pages viewed per iteration
6	MeanPage_P_RFselected	Mean page number the positive feedback images selected from per iteration
7	SelectResultsStrategy	The way of users selecting the search result for completing tasks
8	No_P_SubsetQuery	Number of subset query transition used in positive queries per task
9	No_P_UniqueimagesPerTask	Number of unique images used in positive queries per task
10	No_N_UniqueimagesPerTask	Number of unique images used in negative queries per task
11	PageFindBestResultImage	Page to find the best result image against the ground truth
12	No_P_RepeatQuery	Number of repeat query transition used in positive queries per task
13	TotalPageNRFSelected	Total page number the negative feedback images selected from per task
14	No_Iteration_Query	Number of queries used per task
15	MeanNo_N_ImagesPerQuery	Mean number of image examples used in negative queries per task
16	MaxTimePerIteration	The longest search iteration to complete a task
17	TimePerIterationRange	Time range between the longest search iteration and the shortest search iteration to complete a task
18	TotalNoNRFSelected	Total number of negative feedback images selected per task
19	No_P_OverlapQuery	Number of overlap query transition used in positive queries per task
20	No_N_OverlapQuery	Number of overlap query transition used in negative queries per task
21	TimePerTask	Time used per task
22	TotalNoResultSelected	Total number of result images selected per task
23	No_N_SubsetQuery	Number of subset query transition used in negative queries per task
24	TotalNoPRanking	Total number of positive query image scoring functionality used per task
25	MeanTimePerIteration	Mean time of all search iterations used to complete a task
26	MedianTimePerIteration	Median time of all search iterations used to complete a task
27	TotalNoPageResultViewed	Total number of result pages viewed per task
28	TotalPageSelectedResultFrom	Total page number the result images selected from per task
29	TimeFindBestResultImage	Time when the best result image (against the ground truth) found
30	TotalPagePRFSelected	Total page number the positive feedback images selected from per task
31	No_N_RepeatQuery	Number of repeat query transition used in negative queries per task
32	TotalNoPRFSelected	Total number of positive feedback images selected per task
33	No_N_Q_ImagePerTask	Number of image examples used in negative queries per task
34	TotalNoNRanking	Total number of negative query image scoring functionality used per task
35	No_P_SupersetQuery	Number of superset query transition used in positive queries per task
36	No_P_Q_ImagesPerTask	Number of image examples used in positive queries per task
37	MeanNo_P_ImagesPerQuery	Mean number of image examples used in positive queries per task
38	No_P_QueryImages	Number of images in a positive query per iteration
39	MeanPage_N_RFselected	Mean page number of negative feedback images selected per iteration
40	No_N_SupersetQuery	Number of superset query transition used in negative queries per task
41	No_N_JumpQuery	Number of jump query transition used in negative queries per task
42	TotalNoPHistory	Total number of positive query history functionality used per task
43	TotalNoNHistory	Total number of negative query history functionality used per task
44	MinTimePerIteration	The shortest search iteration to complete a task
45	No_P_JumpQuery	Number of jump query transition used in positive queries per task
46	No_N_RFSelected	Number of negative feedback images selected per iteration
47	No_P_RFSelected	Number of positive feedback images selected per iteration
48	PageN_RFSelected	Total page number the negative feedback images selected from per iteration

Figure 3: 48 unique interaction features generated from the screen capture of the 3 evaluations and the features' descriptions

Interactive feature	Risky / Cautious	Fixed / Evolving	Precise / Weak	Comments
TimePerIteration	Cautious			Based on the definition, cautious users tend to move slowly between two search iterations. Thus, these users likely spend long time in one search iteration.
TotalNoResultSelected			Weak	Based on the definition, users with weak evaluation threshold are lenient on selecting search results, so naturally they will select large number of result images although some images selected as result are only a bit relevant to the query.
TotalNoPHistory		Evolving		Based on the definition, users with evolving information goal are not sure what they are looking for. They like to try different queries and learn from the search result until they have a better idea what they want. As their information goal changes often, they might need to go back to use previous query for better result.

Figure 4: Example of assigning the six characters to the unique interaction features

Predicted interactive features	Regression model	Character
No_PageResultViewed	+ TimePerIteration + No_ResultSelected + MeanPageSelectedFrom* + No_N_QueryImages*	Cautious
No_P_RFSelected	+ MeanPage_P_RFSelected* + No_ResultSelected	Weak
No_ResultSelected	+ TimePerIteration + No_P_RFselected − MeanPageSelectedFrom* + No_PageResultViewed	Cautious; Weak
TotalNoPRanking	+ No_Iteration_Query SelectResultWhileSearching	Precise
TimeFindBestResultImage	− No_Iteration_Query SelectResultWhileSearching − TotalNoPageResultViewed + TimePerTask*	Risky
No_N_JumpQuery	+ No_P_SupersetQuery* + TotalNoPHistory + TotalPageSelectedFrom*	Evolving

Figure 6: Example of assigning the six characters to the regression models