DEIM Forum 2013 A10-5

User Intention in Image Retrieval and Small-Granularity Query Topic Detection

Jiyi LI†, Qiang MA†, Yasuhito ASANO†, and Masatoshi YOSHIKAWA†

† Department of Social Informatics, Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606–8501 Japan

Abstract In image retrieval, for a specific query which can be a keyword or an image sample, different user has different intention and prefers different search results. Although user intention in web page search has been an active research area for several years. There are only a few work on user intention in image retrieval. In this paper, we propose a systematic taxonomy on user intention in image retrieval. User intention can be addressed into two dimensions: "topic" and "intent". Existing work on query "topic" intention in image retrieval use big-granularity topic intention such as "object" and "scene" for a query. However, there are many query topic intention with smaller granularity, such as "black horse" and "zebra" for a "horse" query. We propose a re-ranking approach based on multi-modal relevance feedback to detect small-granularity topic intention and improve the image search results.

Key words Image Retrieval, User Intention

1. Introduction

In image retrieval systems, users can input queries to search and browse images from a database with large amount of images. These queries can be keywords or image samples. Query by keywords refers to Text-based Image Retrieval (TBIR). Query by image samples refers to Content-based Image Retrieval (CBIR). In addition to the image search results, some systems also provide some user interfaces with which users can submit more feedback information on the results. For example, users can label the positive and negative images in the initial results. With these additional relevance feedback information, these systems re-rank the images and provide improved search results to users.

For a specific query in both TBIR and CBIR, different users have different intention and prefer different images in the results. In the relevance feedback session, different users label different instances. The user intention of queries for searching images needs to be considered for image retrieval systems to improve user experience. We therefore investigate existing work for the solutions. However, we find that there are a few work that discuss this issue in image retrieval. In contrast, user intention of queries in web page search has been an active research area for several years. The information in an image and in a web page are different; The motivation of users when search images and when searching web pages cannot establish one-to-one relationship. These lead to the characteristics of image search results and web page search results are different. It results in that we cannot directly use the contributions on user intention in web page search to the topic of user intention in image retrieval.

The research on user intention in web page search can be divided into two classes: "topic" and "intent" [1]. For mining user intention of a given query, on one hand, query topic classification is to identify a query as belonging to one or more categories from a predefined taxonomy. For example, a query of "uefa euro 2012" belongs to a category of "sports/soccer". Some query topic taxonomies have been proposed, e.g. in Ref. [2], Li et al. used 67 categories. On the other hand, query intent classification is to detect the underlying intents of user interactions with search engines. A typical query intent taxonomy is Jansen's three categories: navigational, informational and transactional [3]. For example, navigational intent is to go to a specific website; informational intent is to gather information from one or more web pages. Some approaches has been proposed to detect user intent by mining web search engine log, e.g. [3]. The usage of terms in this area is not unified, some research on query topic classification also use the term of "intent". To clarify the terms in this paper, we denote "intent" as a narrow definition, and "intention" as a general definition including both "topic" and "intent".

In this paper, we give a brief review of the existing work on user intention in web page search as well as in image retrieval. Then we propose a systematic taxonomy on user intention in image retrieval by referring the categories in web page search, on both dimensions of "intent" and "topic". After that, we focus on the issue of how to detect query "topic" in image retrieval. How to detect query "intent" in image retrieval, which is the other dimension of user intention, is not discussed in this paper.

Several existing work on this issue such as Tang et al. [7] use topic intention like "object", "scene", "portrait" and so on for a query. The granularity of these topic intention is big. However, many query topic intention in image retrieval have smaller granularity. For example when a user uses a keyword query of "horse" to search images, the topic intention of this user can be not only "object" or "horse", but also "black horse", "zebra" or "horsemanship". Besides such single concept, topic intention of a query can also be represented by more complex description, such as "a black horse running on a farm". In the search results, different "horse" images have different relevance level because of different topic intention. However traditional image retrieval methods do not consider these detailed query topic intention. When they evaluate the methods, for a "horse" query, all "horse" images are regarded as good results with same relevance level.

We propose a re-ranking approach utilizing visual and textual information of images and user relevance feedback [12] to detect small-granularity topic intention and improve the image search results. We construct a graph model based on the images and related textual information. When users label relevance feedback instances on the initial image search results to represent their query topic intention. We propagate the information including query topic intention through the edges of the graph so that these information can reach the possible target images that the users prefer. Then these target images are upgraded to higher rank positions.

The contributions of this paper are as follows:

- We discuss user intention in image retrieval and propose a systematic taxonomy on both dimensions of query "intent" and query "topic".
- We discuss the granularity of query "topic" intention in image retrieval and propose an approach to detect small-granularity topic intention to improve image search results.

The remainder of this paper is organized as follows. In Section 2, we review the existing work on user intention of web page search. In Section 3, we discuss user intention of image retrieval. In Section 4, we propose an approach which can detect small-granularity query topic intention. In Section 5, we report and discuss the experimental results. We conclude this paper in Section 6.

2. Related Work

In this section, we give a brief review of the existing work on user intention in web page search to illustrate that which issues are discussed on the topic of user intention and concerned by researchers. Ref. [1] provides a detailed survey on this topic. The research in the area can be addressed into two classes: "topic" and "intent" [1].

For mining user intention of a given query, on one hand query topic classification is to identify a query as belonging to one or more categories from a predefined taxonomy. For example, a query of "uefa euro 2012" belongs to a category of "sports/soccer". Some query topic taxonomies have been proposed, e.g. in Ref [2], Li et al. used 67 categories with seven major categories, such as "computer/hardware", "computer/software", "entertainment/music", "living/food and cooking" and so on.

On the other hand, query intent classification is to detect the underlying intents of user interactions with search engines. There are relatively fewer work concentrating on query intent classification than query topic classification in web page search. The state-of-art research divided the underlying intents of user interactions with search engines into three categories [1], [3]: "navigational", "informational" and "transactional". For example, navigational intent is to go to a specific website; informational intent is to gather information from one or more web pages. Jansen et al. [3] found that up to 75% of web queries are single faceted in nature and can be classified into one of the three categories, 80% of the queries are informational in nature.

Several methods have been proposed to automatically classify queries according to their intent. The trends of methods for query intent classification is to mine web search engine log [3], [4]. A state of art method to model the search engine log is query-flow-graph model [5].

3. User Intention in Image Retrieval

In this section, referring the research in web page search, we discuss user intention in image retrieval on both dimensions of query topic classification and query intent classification. In each subsection, we review the existing work in image retrieval first and then discuss how to make a systematic taxonomy for it.

3.1 Query Topic Classification

In contrast to many existing work on user intention in web page search, the work on user intention in image search are relatively fewer. In these work, most of them are actually major in query topic classification if referring the term from user intention in web page search, although they are also named as user intention. For example, Tang et al. [6], [7] proposed approaches based on a taxonomy with topic intention like "object", "scene", "portrait" and so on. We summarize the topic intention they use in Table 1.

These taxonomies with five categories has big-granularity topic intention. However, to further describe and detect user intention, we need topic intention with different granularities, specially with small-granularity.

Table 2 shows two examples with smaller granularity on topic intention. In the first example, for query "apple", the target image of a user may be a "fruit" or a "laptop". The possible target images have different concepts and are from different categories. The second example has smaller granularity than the first example, for query "horse", all these possible target images have the same brief concept of "horse", they are just different on other detailed concepts.

When a user uses a keyword query of "horse" to search images, the topic intention can be "black horse", "zebra" or "horsemanship". Besides such description with single concept, topic intention of a query can also be represented by more complex description, such as "a black horse running on a farm". Traditional image retrieval methods do not consider detailed query topic intention. When they evaluate the methods, for a "horse" query, all "horse" images are regarded as good results.

We divide the granularities of topic intention into three levels, "big", "middle" and "small". Table 3 shows some examples of this three levels. This is not a strict classification. It is to give an intuitive presentation of different granularities. We propose an approach which can detect the small-granularity topic intention as well as middle-granularity and big-granularity topic intention in Section 4.

There are two solutions to meet this requirement. One is to provide results with diverse intents to satisfy as many users as possible. The second is to detect atom user intention dynamically and provide specific results for different users.

3.2 Query Intent Classification

There are only a few work on query intent classification in image retrieval. Because the motivations of users for searching web pages and images are not same, the taxonomies of query intent classification proposed for web page search cannot be used for image retrieval without adaptation [8]~[10]. Lux et al. therefore propose a series of work on this issue.

They first make an exploratory study in Ref. [9] on the relationships between user intents and their browsing behavior including viewed photos and duration of a task. They design ten different search tasks with different intents and ask users to construct queries following these tasks by themselves. The intent taxonomy in this work is still same with the one in web page search. Based on the analysis of the log file, they conclude that only a small part of queries contain explicit intents; user intents may be related to the browsing behavior; the intent taxonomy presented in existing work on web page search is not sufficient for multimedia search.

After that, they propose a work in Ref. [8]. They classify user intents based on browsing behavior using number of viewed photos, number of click results, and duration of a task. They dynamically adapt the results presentation based on different intents. The intent taxonomy they use is navigational, informational-directed, informational-undirected, transactional. The results presentation include list/grid view, ordered by relevance/interestingness, and additional filtering rules. They do not give any technology details on how to classify the intents. When the image results are presented by grid, it is difficult to know how many images have been viewed before the user clicks. There is also no re-ranking process in this work, they only change the results presentation.

Finally, they propose a special taxonomies for image retrieval in Ref. [10]. The disadvantages of this work are as follows: they do not divide TBIR and CBIR which causes that the definitions of categories and category overlaps are ambiguity; the samples they raise do not follow or match their definitions exactly; they only consider single aspect and cannot explain some characteristics of user intents in image retrieval on other aspects, such as range and number of images, modality of query and information need, semantic relevance degree between target and query, and so on.

For example, in their taxonomy, their define "mental image" as matching the retrieved images to the mental image. We think that their "mental image" is CBIR because they announce that: it is by visual analysis and image semantics; there is not overlap between "mental image" and "navigation" because a search goal is either to find the image by content or by semantics. But their sample for a category overlap based on "mental Image" is "searching for a background image for your cell phone" which is not CBIR. Another example is that they define "navigation" which is TBIR as "the user knows about the existence of the image, but its content is unknown". They also define the case of "the user knows how the image content looks like" as belonging to "Mental Image" which is CBIR. Therefore, there is no category to describe the case of "the user knows how the image content looks like and uses TBIR to search it".

We propose a systematic taxonomy considering query modality and characteristics of image retrieval. Table 4 and 5 illustrate the taxonomy in the case of TBIR and CBIR respectively.

This taxonomy is approach-oriented. It means that different classes should be solved with different approaches. Therefore there is no categories like "transaction" in web

Table 1 Existing Work on Query Topic Classification in Image Retrieval

Ref.	Query Topic Intention	Scenario
[6]	Scene, Object, People, Portrait, General	CBIR
[7]	General Object, Object with simple background, Scene, Portrait, People	TBIR, Query Expansion

Table 2 Examples for Query Topic Classification in Image Retrieval

Query 1	Possible Target Images			
apple				
Topic	food/fruit	computer/laptop	building/shop	multi-categories

Query 2	Possible Target Images				
horse					
Topic	black horse	zebra	on farm	horsemanship	horses

Table 3 Topic Intention on Different Granularities

Level	Description	Example
Big	Image Property	Scene, Object, Portrait
Middle	Brief Description	Fruit, Laptop, Shop
Small	Detailed Description	Black Horse, on farm, horsemanship

Table 4 Query Intent Classification for TBIR

Class	Description	Example
Single Target	I want to search a specific image	mascot of Olympics 2012 the cd cover and booklet of Mendelssohn's Violin Concerto in E minor by Itzhak Perlman
Simple Collection	I want to search some images in one facet or one topic	Kiyomizu Temple Ipad2 Michael Jordan
Complex Collection	I want to search some images in multi-facets or multi-topics or for a task	Tour in Kyoto (images of various tourist attractions in Kyoto) Apple Products Famous Basketball players
Multimedia Information	I want to search the images and related description	Tour in Kyoto (image collections, travel notes, travel route recommendation, hotel) Ipad2 (images, product introductions, comments, prices in stores) Michale Jordan (images, biography, news, interview)

page search which means users want to download images. Users search the images they want, whether download or not is based on their further purpose. It is independent with other user intents and the usage of search approaches.

4. Small-Granularity Query Topic Intention Detection

In section 3, we have discussed user intention in image retrieval. In this section, we focus on the issue of how to detect query "topic". The other dimension of user intention, how to detect query "intent", is not yet discussed in this paper. We utilize the approach we propose in [12] to detect the small-granularity query topic intention. Note that our approach can also handle middle granularity and big granularity.

We propose a re-ranking approach utilizing visual and textual information of images and user relevance feedback [12] to detect topic intention on different granularities and improve the image search results. We construct a graph model based on the images and related textual information. When users label relevance feedback instances on the initial image

Table 5 Query Intent Classification for CBIR

Class	Description	Example
Semantic Consistency	I want to search some image which are visual similar to the query image	Query: a "horse" image Target: "horse" images
Semantic Relativity	I want to search some images which are not visual similar to the query image but related to the query image	Query: a "horse" image Target: some "zebra" images some "cow" images some "farm" images some "horsemanship" images

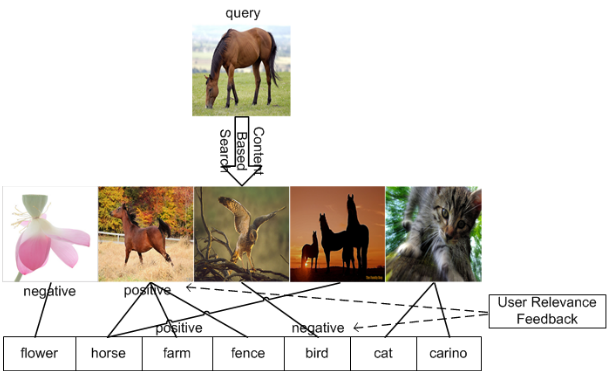

Figure 1 Scenario for the Discussion of Query Topic Intention

search results to represent their query topic intention. We propagate the information including query topic intention through the edges of the graph so that these information can reach the possible target images that users prefer. Then these target images are upgraded to higher rank positions.

4.1 Formulation

Without loss of generality, we define and formulate the following scenario for this discussion. We utilize the images from Flikcr which is a social image hosting websites on which users can upload, tag and share their images. The textual information we utilize here is the social tags of images. We define a scenario with Flickr, CBIR and social tag. The case of other image sources, TBIR, other textual information can also be formulated in a similar way.

For a given query image q, a content-based image retrieval method computes the top-n content-based similar image results $\mathcal{A} = \{a_1, \cdots, a_n\}$ from a social image database $\mathcal{D}$. Let s_{iq} be the similarity between q and a_i. We regard $\mathcal{A}$ as the candidate image set and the social tags $\mathcal{T} = \{t_1, \cdots, t_m\}$ of images in $\mathcal{A}$ as the candidate tag set. We define $\mathcal{T}_{a_i}$ as the tag set of each image $a_i \in \mathcal{A}$. Users can select relevance feedback instances from $\mathcal{A}$ and $\mathcal{T}$. We define $\mathcal{RF_A}$ and $\mathcal{RF_T}$ as the social image and tag relevance feedback instance sets. Our task is to re-rank the image set $\mathcal{A}$ with both the tag set $\mathcal{T}$ and the relevance feedback set $\mathcal{RF_A}$ and $\mathcal{RF_T}$. Figure 1 illustrates our definition.

4.2 Image-Tag Relationship Model

To leverage both visual information of social image and textual information of social tag for re-ranking, we construct a graph model, shown in Figure 2, with the candidate sets $\mathcal{A}$ and $\mathcal{T}$ for analyzing image-tag relationships. The vertices of the graph model denote social images and their tags. Note that query image q contains no textual information.

The edges of the graph model denote the relationships among images and tags. There are three kinds of image-tag relationships: image-to-image relationship, tag-to-tag relationship, and image-to-tag annotation relationship. The first two kinds of relationships reflect the intra-relationships among images or tags. The third relationship reflects the inter relationship between images and tags.

4.3 Our Propagation-Based Approach

Following the image-tag annotation relationships in the graph model, we propagate the rank scores of images in $\mathcal{A}$ and tags in $\mathcal{T}$ along the links between images and tags. We observe a phenomenon that for an image a_i, when propagating the rank scores from images to tags, if a_i has a high rank score, its related tags will obtain higher rank scores. When propagating the rank scores is from tags to images, if the related tags of a_i have high rank scores, a_i will obtain a higher rank score. On the other hand, for a tag t_x, it also has similar phenomenon. Therefore, we naturally arrive at the following mutual reinforcement assumption: a high-ranked image for q is one to which many high-ranked tags point; a high-ranked tag for q is a tag that points to many high-ranked images. The iterative formulas for computing the rank scores are defined as follows:

Initialization:

$$\mathcal{Q}'_0(a_i) = \Phi(vd_i),\ \ \mathcal{Q}'_0(t_x) = \Phi(td_x);\ \ 0 \leqq \alpha, \beta \leqq 1$$

Iteration:

$$\begin{cases} \mathcal{Q}_{k+1}(t_x) = \alpha\Phi(td_x) + (1-\alpha)\sum_{\forall a_i : t_x \in T_{a_i}} \Phi(vd_i)\mathcal{Q}'_k(a_i) \\ \mathcal{Q}_{k+1}(a_i) = \beta\Phi(vd_i) + (1-\beta)\sum_{\forall t_x : t_x \in T_{a_i}} \Phi(td_x)\mathcal{Q}'_k(t_x) \\ \mathcal{Q}'_{k+1}(t_x) = \Phi(\mathcal{Q}_{k+1}(t_x)),\ \ \mathcal{Q}'_{k+1}(a_i) = \Phi(\mathcal{Q}_{k+1}(a_i)) \end{cases}$$

$$\Phi(\mathcal{Q}_k(a_i)) = \frac{Q_k(a_i) - min_j\{Q_k(a_j)\}}{max_j\{Q_k(a_j)\} - min_j\{Q_k(a_j)\}},$$

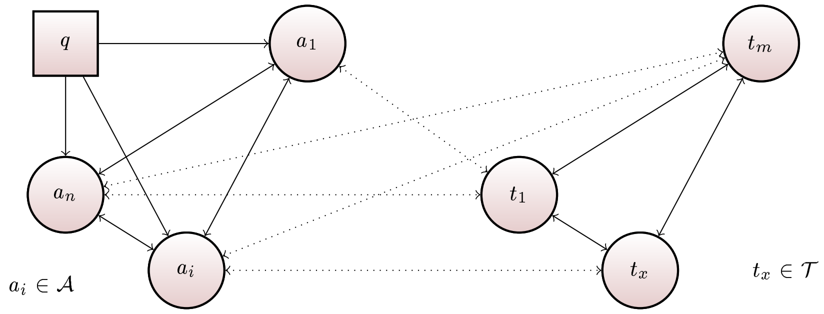

Figure 2 Image-Tag Relationship Model

$$\Phi(\mathcal{Q}_k(t_x)) = \frac{Q_k(t_x) - min_y\{Q_k(t_y)\}}{max_y\{Q_k(t_y)\} - min_y\{Q_k(t_y)\}}$$

The iteration parameters α and β are damping factors. k is the number of iteration steps. We initiate $\mathcal{Q}'_0(t)$ of tags with textual descriptors and $\mathcal{Q}'_0(a)$ of images with visual descriptors. $\mathcal{Q}'_k(\cdot)$ is the normalized rank score of $\mathcal{Q}_k(\cdot)$ by using min-max normalization method.

For a candidate image a_i, its image similarity to the query image is an inherent property. Images that have high similarity can be regarded as more important on the graph. For a candidate tag t_x, it is also similar. Therefore we use visual descriptors and textual descriptors as the weights in the iterations. These weights represent the importance of these images and tags on the graph.

We develop a series of rules to add to the basic mutual reinforcement process for utilizing multi-modal relevance feedback information. The graph model allows such modifications easy to implement. The key idea is that we mark relevance feedback instances on the graph, upgrade (downgrade) the score of positive (negative) instances, and propagate these scores through the links between images and tags, so that we can refine the rank scores of the related images and tags using the relevance feedback information. After several iterations, the relevance feedback information is propagated to all other related images and tags.

All these rules enforcedly change the value of some terms in the iteration formulas of the basic mutual reinforcement process, while the form of the iteration formulas are not changed. In the following rules, we use "+" to denote positive instances and "-" to denote negative instances.

Rule 1: At the beginning of each round, we use the rank results of the last round to initialize the rank scores.

$$\mathcal{Q}_0^{l+1} = \mathcal{Q}_h^l,$$

where l is the round number and h is the iteration number of the previous round.

Rule 2: At the beginning of each round, we change the value of the visual and textual descriptors of each relevance feedback instance.

$$vd_+ = \max_i\{vd_i\},\ vd_- = \min_i\{vd_i\},$$

$$td_+ = \max_x\{td_x\},\ td_- = \min_x\{td_x\}$$

Rule 3: We don't compute the rank scores of relevance feedback instances. They are fixed to current maximum (minimum) scores in the candidate set.

$$\mathcal{Q}_k(a_+) = \max_i \mathcal{Q}_k(a_i),\ \mathcal{Q}_k(a_-) = \min_i \mathcal{Q}_k(a_i),$$

$$\mathcal{Q}_k(t_+) = \max_x \mathcal{Q}_k(t_x),\ \mathcal{Q}_k(t_-) = \min_x \mathcal{Q}_k(t_x),\ 0 \leqq k \leqq h$$

With these rules, positive instances can contribute more scores in the propagation and refining process so that their related images and tags can gather more scores and be ranked higher; negative instances can only contribute fewer scores so that their related ones can gather fewer scores and be ranked lower.

4.4 Small-Granularity Detection

In this subsection, we describe how our re-ranking approach handle the issue of query topic intention in social image retrieval on small topic granularity.

Figure 3 illustrates an example. Following the "horse" query in Table 2, we add another "horse" image to describe this issue. When a user labels a "black horse" image as positive instance, his topic intention is to search for a "black horse". Assume that the images has been well tagged, and these two "black horse" images in this example has been both tagged with "black". Then with our approach, in contrast to other "horse" images, additional information are propagated through the links of positive labeled "black horse" image, "black" tag and unlabeled "black horse" image. By this way, the unlabeled "black horse" image can gather more scores during the propagation and iterations, and so that it can be ranked higher than other "horse" images which are not "black horse". Our approach can also handle topic intention with middle- and big-granularity by the same way.

Furthermore, our approach can not only handle the case of topic intention description with single concept, but also handle the case with complex description, e.g. "a black horse running on farm". In such case, the user intention of the

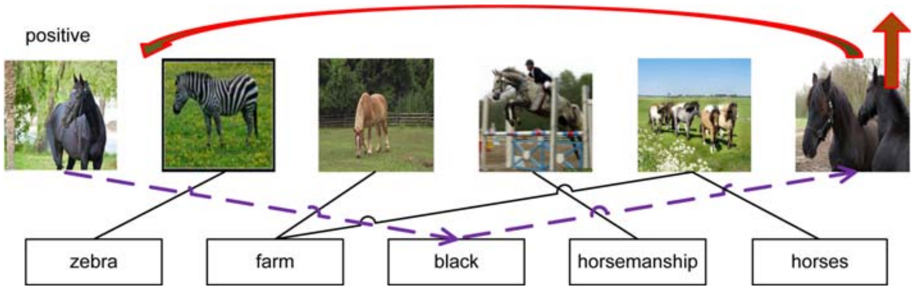

Figure 3 Query Topic Intention in Social Image Re-ranking

positive instances is propagated through the edges of the graph and the nodes of tags "black", "running" and "farm". The images which are related to one or more tags in these tags will be ranked higher than other images. The images related three of these tags will be ranked higher than the images related only one or two of these tags.

5. Experiment

5.1 Experimental Settings

The dataset we used is NUS-WIDE [13]. It was created by downloading images and their social tags from Flickr. It has 269,648 images and about 425,000 unique original tags. For images, it provides six types of low-level features extracted from the images. For tags, the authors of this dataset set several rules to filter the original tag set. They delete those tags that have a frequency below a certain threshold; the low frequency threshold is set to 100. They also remove the tags that do not exist in WordNet. At the end, they counted 5,018 unique tags. We retain this filtering in our experiment because it reduces the noises in the tag set.

The evaluation metric used in our experiment is Normalized Discounted Cumulative Gain (NDCG) [14]:

$$NDCG@k = Z_k \sum_{j=1}^{k} ((2^{r(j)} - 1)/log(1 + j)).$$

It is an effective metric for evaluating the rank results with relevance levels. $r(j)$ is the relevance level of the image at rank j. Z_k is a normalization constant and equal to the maximum DCG value that the top-k ranked images can achieve, such that the NDCG score is equal to 1 for optimal results.

5.2 Traditional Evaluation Method and Middle-Granularity Evaluation

In traditional evaluation methods for image retrieval do not consider detailed query topic intention. When they evaluate the methods, for a "horse" query, all "horse" images are regarded as good results.

The images in the image results are labeled with five relevance levels by us, according to their visual and semantic relevance to the query images. The range of relevance levels is from 0 to 4: irrelevant (0), weakly relevant (1), partially relevant (2), relevant (3), and very relevant (4). Table 6 shows an intuitive example of different relevance levels.

NUS-WIDE provides image annotation ground-truth of 81 concepts for the entire dataset, but it does not appoint a query sample set or provide ground-truth for content-based image retrieval. We need to construct these components by ourselves for our experiments. In our experiments, we randomly choose 100 images as a query image set (from the entire dataset) for our evaluation. We choose 20 queries in this query image set as a training set; the other 80 queries become the testing set. Note that there is no textual information available for these images when they are used as queries. For each query, we re-rank the images in top-n content-based similar image results, the cut-off size $n = 100$.

This evaluation can evaluate the approaches on middle-granularity topic intention such as "horse". In Ref. [12], a series of experiments have illustrated that our approach has better performance than other approaches on this evaluation.

5.3 Evaluation with Query Topic Detection and Small-Granularity Evaluation

The experiments for this section is still work-in-progress. We design the experiment in this section. Traditional evaluation methods cannot reflect query topic intention of atom user and therefore cannot be used here. We need to design special evaluation method for atom user intention.

Sakai et al. propose an evaluation method for evaluating user intention [11]. However, their method cannot be utilized here. One reasons is that they focus on navigational and informational intents which refer to query intent detection. The other reason is that their method is proposed for evaluating the diversity. An approach performs better if its results can cover more potential intention of different users. In our case, we evaluate whether an approach can generate results for atom users. It means that for different users, it needs to generate different results.

We use TBIR for this evaluation. We first select 20 concepts on middle-granularity level to construct the queries, e.g., "horse", "apple", "sky" and so on. For each query , we set 5 different cases to simulate different small-granularity topic intention. In each case we select a concept on small-

Table 6 Sample of Relevance Levels for NDCG in Traditional Evaluation Method

Query	Image Samples				
					
Relevance Level	4	3	2	1	0

Table 7 Sample of Levels for NDCG in Small-Granularity Topic Intention Detection

Query	Topic Intention	Image Samples					
horse	black horse						
NDCG Level		5	4	3	2	1	0

granularity level, e.g., "black", "zebra", "horsemanship", and so on. In total, we get 100 test cases.

We still use the metric of NDCG for evaluation. However, the rule of labeling the results is different. The images in the image results are labeled with six levels by us, according to their semantic relevance to the query images and the topic intention in the test case. The range of relevance levels is from 0 to 5: irrelevant (0), weakly relevant (1), partially relevant (2), relevant (3), very relevant (4), and very relevant and match the topic intention (5). Table 7 shows an intuitive example of different levels.

6. Conclusion

In this paper, we provide a discussion on user intention in image retrieval by referring the existing work on user intention in web page search. On the other hand, we focus on query topic intention in image retrieval with small-granularity while existing work on query topic intention use big-granularity topic intention. We propose a re-ranking approach based on multi-modal relevance feedback to detect small-granularity topic intention and improve the image search results.

References

[1] Brenes, David J. and Gayo-Avello, Daniel and Pérez-González, Kilian: "Survey and evaluation of query intent detection methods", Proceedings of the 2009 workshop on Web Search Click Data (WSCD'09), pp.1–7, 2009

[2] Li, Ying and Zheng, Zijian and Dai, Honghua: "KDD CUP-2005 Report: Facing a Great Challenge", ACM SIGKDD Explorations Newsletter, Vol.7, No.2, pp. 91-99, 2005

[3] Jansen, Bernard J. and Booth, Danielle L. and Spink, Amanda: "Determining the informational, navigational, and transactional intent of Web queries", Information Processing and management, Vol.44, Iss.3, pp.1251–1266, 2008

[4] Rose, Daniel E. and Levinson, Danny: "Understanding User Goals in Web Search", Proceedings of the 13th international conference on World Wide Web (WWW'04), pp.13–19, 2004

[5] Boldi, Paolo and Bonchi, Francesco and Castillo, Carlos and Donato, Debora and Gionis, Aristides and Vigna, Sebastiano: "The query-flow graph: model and applications", Proceedings of the 17th ACM conference on Information and knowledge management (CIKM'08), pp. 609–618, 2008

[6] Cui, Jingyu and Wen, Fang and Tang. Xiaoou: "User intention modeling for interactive image retrieval", 2010 IEEE International Conference on Multimedia and Expo (ICME'10), pp.1517–1522, 2010

[7] Tang, Xiaoou and Liu, Ke and Cui, Jingyu and Wen, Fang and Wang, Xiaogang: "IntentSearch: Capturing User Intention for One-Click Internet Image Search", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Vol.34, Iss.7, pp.1342–1353, 2012

[8] Lux, Mathias and Kofler, Christoph, "Dynamic Presentation Adaptation based on User Intent Classification", Proceedings of the 17th ACM international conference on Multimedia (MM'09), pp.1117–1118, 2009

[9] Kofler, Christoph and Lux, Mathias, "An Exploratory Study on the Explicitness of User Intentions in Digital Photo Retrieval", Proceedings of I-KNOW '09 and I-SEMANTICS '09, pp.208–214, 2009

[10] Lux, Mathias and Kofler, Christoph and Marques, Oge: "A classification scheme for user intentions in image search", Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (CHI EA '10), pp. 3913–3918, 2010

[11] Sakai, Tetsuya: "Evaluation with Informational and Navigational Intents", Proceedings of the 22nd international conference on World Wide Web (WWW'12), pp.13–19, 2012

[12] Li, Jiyi and Ma, Qiang and Asano, Yasuhito and Yoshikawa, Masatoshi: "Re-ranking by Multi-Modal Relevance Feedback for Content-Based Social Image Retrieval", Proceedings of the 14th Asia-Pacific Web Conference (APWeb'12), pp.399-410, 2012

[13] Chua, Tat-Seng and Tang, Jinhui and Hong, Richang and Li, Haojie and Luo, Zhiping and Zheng, Yantao: "NUS-WIDE: A Real-World Web Image Database from National University of Singapore", Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR'09), pp.48:1–48:9, 2009

[14] Jrvelin, Kalervo and Keklinen, Jaana: "Cumulated gain-based evaluation of IR techniques", ACM Transactions on Information Systems (TOIS), Vol.20, Iss.4, 2001