

Inferring User Image-Search Goals Under the Implicit Guidance of Users

Zheng Lu, Xiaokang Yang, *Senior Member, IEEE*, Weiyao Lin, Hongyuan Zha, and Xiaolin Chen

Abstract—The analysis of user search goals for a query can be very useful in improving search engine relevance and user experience. Although the research on inferring user goals or intents for text search has received much attention, little has been proposed for image search. In this paper, we propose to leverage click session information, which indicates high correlations among the clicked images in a session in user click-through logs, and combine it with the clicked images’ visual information for inferring user image-search goals. Since the click session information can serve as past users’ implicit guidance for clustering the images, more precise user search goals can be obtained. Two strategies are proposed to combine image visual information with the click session information. Furthermore, a classification risk based approach is also proposed for automatically selecting the optimal number of search goals for a query. Experimental results based on a popular commercial search engine demonstrate the effectiveness of the proposed method.

Index Terms—Click-through logs, goal images, image-search goals, semi-supervised clustering, spectral clustering.

I. INTRODUCTION

IN WEB SEARCH applications, users submit queries (i.e., some keywords) to search engines to represent their search goals. However, in many cases, queries may not exactly represent what they want since the keywords may be polysemous or cover a broad topic and users tend to formulate short queries rather than to take the trouble of constructing long and carefully stated ones [1]–[3]. Besides, even for the same query, users may have different search goals. Fig. 1 shows some example user image-search goals discussed in this paper. Each goal in Fig. 1 is represented by an image example. From Fig. 1 and our experimental results, we find that users have different search goals for the same query due to the following three reasons.

Manuscript received June 3, 2012; revised September 29, 2012 and January 15, 2013; accepted March 15, 2013. Date of publication March 28, 2013; date of current version March 4, 2014. This work was supported by the NSFC (61025005, 61129001, 61221001 and 61201446), the 973 Program (2010CB731401 and 2010CB731406), the 111 Project (B07022), and the Open Project Program of the National Laboratory of Pattern Recognition and NSF under Grant IIS-1049694. This paper was recommended by Associate Editor R. Hamzaoui.

Z. Lu, X. Yang, W. Lin, and X. Chen are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xkyang@sjtu.edu.cn; luzheng@gmail.com; hellomikelin@gmail.com; xiaolin.chen317@gmail.com).

H. Zha is with the College of Computing, Georgia Institute of Technology, Atlanta, GA USA (e-mail: zha@cc.gatech.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCSVT.2013.2255418

- 1) Multi-concepts: a keyword may represent different things. For example, besides being a kind of fruit, “apple” is endowed with new concepts by Apple, Inc.
- 2) Multi-forms: the same thing may have different forms. Take “Bumblebee” in the film *Transformers* as an example. It has two modes: car mode and humanoid mode. These two modes are the two forms of “Bumblebee.”
- 3) Multi-representations: in image search, the same thing can be represented from different angles of view such as the query leaf. It can be represented in a real scene or by a close-up.

Inferring user search goals is very important in improving search-engine relevance and user experience. Normally, the captured user image-search goals can be utilized in many applications. For example, we can take user image-search goals as visual query suggestions [4] to help users reformulate their queries during image search. Besides, we can also categorize search results [5] for image search according to the inferred user image-search goals to make it easier for users to browse. Furthermore, we can also diversify and re-rank the results retrieved for a query [6], [7] in image search with the discovered user image-search goals. Thus, inferring user image-search goals is one of the key techniques in improving users’ search experience.

However, although there has been much research for text search [8]–[12], few methods were proposed to infer user search goals in image search [4], [13]. Some works try to discover user image-search goals based on textual information (e.g., external texts including the file name of the image file, the URL of the image, the title of the web page that contains that image and the surrounding texts in image search results [14] and the tags given by users [4]). However, since external texts are not always reliable (i.e., not guaranteed to precisely describe the image contents) and tags are not always available (i.e., the images may not have corresponding tags that need to be intentionally created by users), these textual information based methods still have limitations.

It should be possible to infer user image-search goals with the visual information of images (i.e., image features) since different image-search goals usually have particular visual patterns to be distinguished from each other. However, since there are semantic gaps [15] between the existing image features and the image semantics, inferring user image-search goals by visual information is still a big challenge. Therefore, in this paper, we propose to introduce additional information sources to help narrow these semantic gaps.



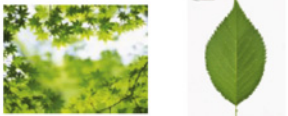
Query	Different user image-search goals
1. apple	
2. Bumblebee	
3. leaf	

Fig. 1. Different user image-search goals represented by image examples in image search by our experiment.

Intuitively, the click-through information from the past users can provide good guidance about the semantic correlation among images. By mining the user click-through logs, we can obtain two kinds of information: the click content information (i.e., the visual information of the clicked images) and the click session information (i.e., the correlation information among the images in a session). Commonly, a session [16] in user click-through logs is a sequence of queries and a series of clicks by the user toward addressing a single information need. In this paper, we define a session in image search as a single query and a series of clicked images as illustrated in Fig. 2. Usually, the clicked images in a session have high correlations. This correlation information provides hints on which images belong to the same search goal from the viewpoint of image semantics. Therefore, in this paper, we propose to introduce this correlation information (named as click session information in this paper) to reduce the semantic gaps between the existing image features and the image semantics. More specifically, we propose to cluster the clicked images for a query in user click-through logs under the guidance of click session information to infer user image-search goals. With the introduction of the correlation information, the reliability of visual features can be improved.

The contributions in this paper can be described as follows.

- 1) We propose a new framework that combines image visual information and click session information for inferring user image-search goals for a query. In this way, more precise image-search goals can be achieved.
- 2) We propose two strategies (i.e., the edge-reconstruction-based strategy and the goal-image-based strategy) to effectively implement the process of combining image visual information with click session information. We also propose to introduce spectral clustering for handling the arbitrary cluster shape scenario during clustering.
- 3) Since different queries may have different number of search goals (e.g., some queries may have two goals while others may have three goals as in Fig. 1), we further propose a classification risk (CR)-based approach

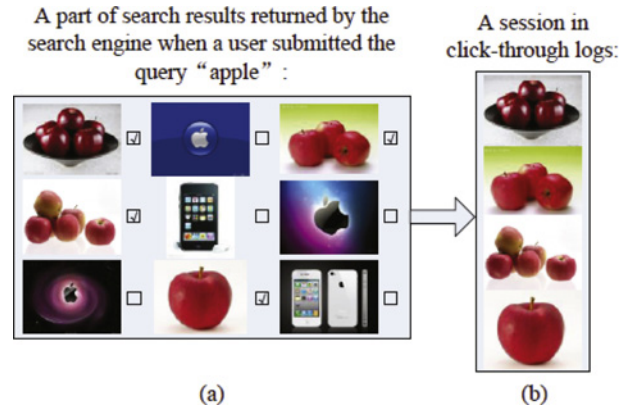


Fig. 2. Session for the query apple in user click-through logs. (a) Search results returned by the search engine. The check marks mean that the images were clicked by the user. (b) Session in user click-through logs.

to automatically decide the optimal number of search goals for a query.

The rest of this paper is organized as follows. Section II introduces some related works. The framework of our approach is described in Section III. Section IV introduces the edge-reconstruction-based strategy to combine image visual information with click session information and Section V introduces the goal-image-based strategy. Section VI describes the clustering method for achieving search goals as well as the CR-based approach to optimize the number of user search goals. The experimental results are given in Section VII. Section VIII concludes the paper.

II. RELATED WORK

In recent years, the research on inferring user goals or intents for text search has received much attention [8]–[11], [17]. Many early researches define user intents as navigational and informational, [8], [9] or by some specific predefined aspects, such as product intent and job intent [10]. Some works focus on tagging queries with more hierarchical predefined concepts to improve feature representation of queries [17]. However, in fact, these applications belong to query classification. User search goals and the number of them should be arbitrary and not predefined. Some works analyze the clicked documents (i.e., click content information) for a query in user click-through logs to explore user goals [11]. However, the click session information is not fully utilized.

Although there has been much research on inferring user goals for text search, few methods were proposed in image search [4], [13]. Zha *et al.* [4] try to capture user goals to give visual suggestions for a query in image search. They first select some tag words as textual suggestions by satisfying two properties: relatedness and informativeness. Then, they collect the images associated with a suggested keyword and cluster these images to select representative images for the keyword. However, the good performance of their method depends on the precision of tags (i.e., tags that are manually created by users, such as the tags in Flickr [18]). In many web image search engines, manual tags are not available and only external

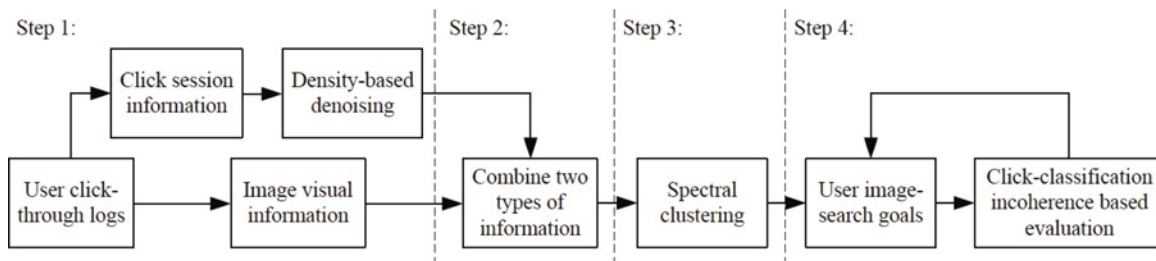


Fig. 3. Framework of our approach.

texts are achievable (e.g., Baidu image [19] and Google image [20]). In these cases, the performance of [4] may be decreased by using external texts as the external texts are not as reliable as tags.

The research on diversity in retrieval [6], [7], [21], [22] is relevant to user goal inference. It aims to diversify the results retrieved for an ambiguous query, with the hope that at least one of the interpretations of the query intent will satisfy the user. In early works, Carbonell *et al.* [21] introduced marginal relevance into text retrieval by combining query-relevance with information-novelty. This information-novelty can be considered as low-level textual content novelty. Recent works [6], [7] model the diversity based on a set of sub-queries. The sub-queries are generated by simply clustering the documents in search results or by query expansion. This diversity can be considered as high-level semantic diversity. The research on diversity in image retrieval has just started [22]. We consider the diversity and novelty of image retrieval as high-level image semantic diversity and low-level visual content novelty, respectively. The inferred user image-search goals in this paper can exactly be utilized to diversify the image search results from high-level image semantics.

Our goal-inference method is based on image clustering using similarity graphs. There has been some research on image clustering with different types of information [14], [23]. Cai *et al.* [14] first use textual and link information to cluster the images in web pages, and then they use visual information to further cluster the images in each cluster. They consider that a single web page often contains multiple semantics and the blocks in a page containing different semantics (instead of pages) should be regarded as information units to be analyzed. They define link information as the relationships between page, block, and image. However, when we cluster the images for a query to infer user goals, there are no such blocks or link information. Instead, we use click information in this paper. Cheng *et al.* [23] first divide a session into the positive part ξ^+ and the negative part ξ^- . After that, they merge the positive parts into chunklets only if the positive parts contain an image in common, and the edges between chunklets are then added if the images in ξ^+ and ξ^- of a session appear in two chunklets, respectively. Finally, clustering is implemented on the chunklet graph. Although their method tried to introduce user information for facilitating visual information, it still has limitations since this method requires the users to identify ξ^+ and ξ^- in each session. However, in real data, it is difficult to divide ξ^+ and ξ^- precisely and ensure that the images in a chunklet will not appear in both ξ^+ and ξ^- of a session

simultaneously. Poblete *et al.* [24] propose to use queries to reduce the semantic gap. They define the semantic similarity graph as an undirected bipartite graph, whose edges connect a set of relative queries and the clicked images for these queries. However, if the set of queries are irrelative, there may be few or no images shared by multiple queries (e.g., the users submitting the different queries do not click the same image). In this case, the queries and their clicked images in the bipartite graph are independent and the semantic similarity graph cannot provide any semantic information. This situation often happens if we randomly select a small set of queries from query logs (i.e., do not purposely select the specific relative queries). Comparatively, in this paper, we use the clicks by different users for the same query to reduce the semantic gap. Thus, our algorithm is flexible to construct the semantic similarity graph for an individual query instead of a set of queries.

III. FRAMEWORK OF OUR APPROACH

The framework of our proposed user image-search goal inference method is shown in Fig. 3. Our framework includes four steps.

1) *Step 1*: We first extract the visual information of the clicked images from user click-through logs. Normally, the images clicked by the users with the same search goal should have some common visual patterns, while the images clicked by the users with different search goals should have different visual patterns to be distinguished from each other. For example, for the query apple, there must be some visual patterns to distinguish fruit apples from phones. Therefore, it is intuitive and reasonable to infer user image-search goals by clustering all users' clicked images for a query with image visual information and use each cluster to represent one search goal. In this paper, we extract three types of image visual features (i.e., color, texture, and shape features) containing color moments (CMG) [16], color correlogram (CC) [6], co-occurrence texture (CT) [4], local binary pattern (LBP) [14], and edge auto-correlogram (EAC) [13]. We concatenate the above five feature channels to get the feature vector for each image.

At the same time, we also extract the click session information from user click-through logs. We consider that the clicked images in a session have high correlations, which is under the hypothesis that the user has only one search goal when he submits a query and he just clicks those similar images. However, in the real situation, many users may click some noisy images. For example, even if a user only wants to search

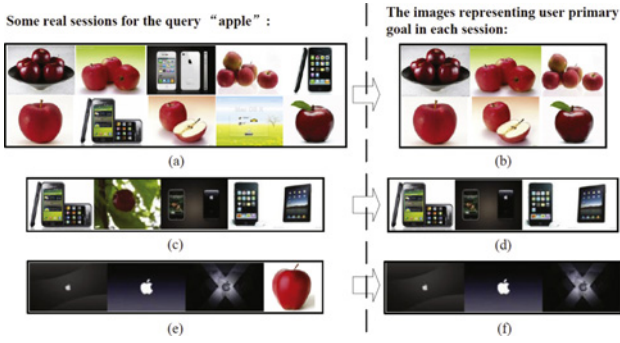


Fig. 4. Some real sessions for the query “apple” in our click-through logs. The right part shows the images picked out to represent user primary goal in each session.

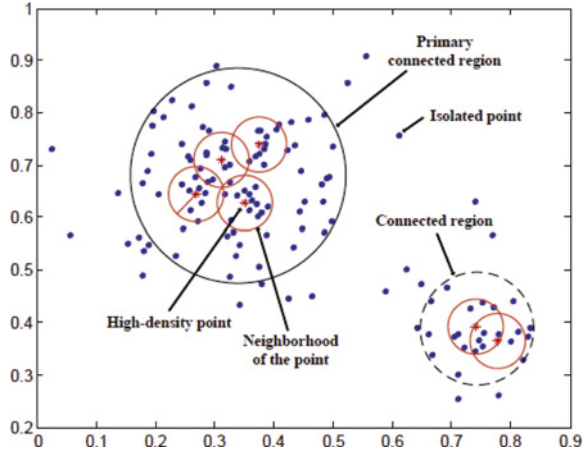


Fig. 5. Density-based method to pick off the isolated images and pick out the images that represent user primary goal in a session.

the fruit apple, at the beginning when he submits the query apple, he may also click some images about iPhone and even click some completely irrelevant images by mistake as shown in Fig. 4(a). If these noisy images are included, the click session information will become less meaningful. In order to solve this problem, a denoising process needs to be performed such that the images representing user primary goal in a session can be picked out to reflect click session information as shown in Fig. 4(b), and the other images in the session are all considered as noises and will be deleted in the session. Fig. 4(c)–(f) show another two examples. Intuitively, three steps should be implemented: picking off the isolated images, clustering the images into clusters that represent potential user goals, and choosing the biggest cluster that has the most images to represent user’s primary goal. In this paper, we propose a density-based method instead of the above three steps to select the images that represent user primary goal. In Fig. 5, each point represents an image in a session, the smallest circle represents the neighborhood of the point, the dashed circle is the connected region, and the biggest circle is the primary connected region. The distance between two points is defined as follows:

$$Dis_{ij} = 1 - \cos(I_i, I_j) = 1 - \frac{I_i \cdot I_j}{|I_i||I_j|} \quad (1)$$

where I_i is the normalized feature vector of the i th clicked image. The density of a point is defined as the number of the points in the neighborhood of it. The radius of the neighborhood normally needs to be set properly. If the radius is too large, the de-noising effect will be not obvious. If the radius is too small, the session will contain less information. In our experiment, we set the radius to be 0.1 according to the experimental statistics. We find some points with the highest densities and merge the neighborhoods of them into a connected region if they have most of the points in common (90% in this paper). We repeat this process until the connected regions cannot be merged again. Finally, we choose the biggest connected region, which has the most points, to represent user’s primary goal. Thus, these similar images, instead of all the images in the session are used to provide click session information.

2) *Step 2*: Image visual information is combined with click session information for further clustering by one of the two proposed strategies, named edge-reconstruction-based strategy and goal-image-based strategy. It should be noted that these two strategies are alternatives by using different ways to model the clicked images for a query with similarity graph [25]. The edge-reconstruction-based strategy utilizes click session information to reconstruct the edges in the similarity graph, while the goal-image-based strategy utilizes click session information to represent the vertices. The basic ideas of the two proposed strategies are shown in Fig. 6. In Fig. 6, the smaller points (including the circle points and the star points, which represent the ground truth of two clusters) represent the clicked images projected into a 2-D space according to their feature vectors in Step 1. The dashed ellipses represent clustering results. If we only use image visual information to cluster the images, the clustering result may be far from satisfactory, as in Fig. 6(a). Therefore, we propose to use click session information for helping clustering. As mentioned, since the click-through logs may indicate relationships among images [e.g., images belong to the same session imply their high correlation as the solid-line ellipses in Fig. 6(b)], by including click session information, more precise clustering can be achieved. Basically, our proposed edge-reconstruction-based strategy utilizes click session information as semi-supervised information to modify the mutual connectivity between the images in the similarity graph, as in Fig. 6(c). Thus, a more reasonable connectivity graph can be achieved and the clustering results can be improved. The second proposed goal-image-based strategy tries to fuse the images in the same session into a super-image [named as goal image in this paper, i.e., the bigger point in Fig. 6(d)] and the search goals are inferred by clustering these goal images rather than the original images. This process can be viewed as a re-sampling process that re-samples the original images in a session into a goal image. We will describe these two strategies in detail in Sections IV and V, respectively.

3) *Step 3*: We propose to introduce spectral clustering algorithm [25] to cluster the image graph that contains both image visual information and click session information. Spectral clustering is introduced in this step because clusters representing different user goals may have arbitrary shapes in visual feature

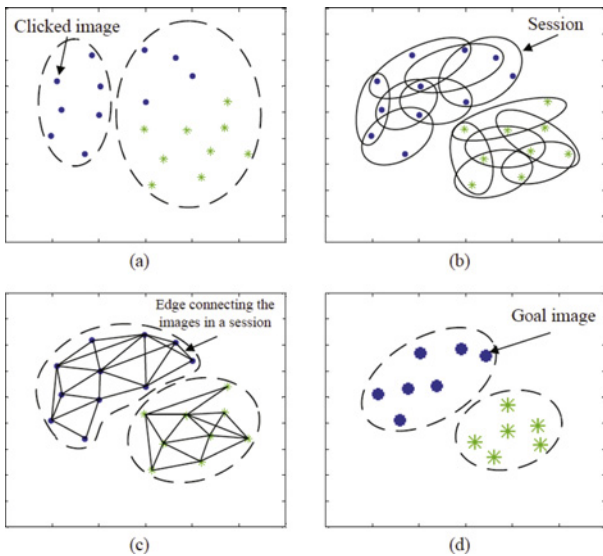


Fig. 6. Clustering the clicked images with the help of click session information. The smaller points represent the clicked points. The circle and star points represent the ground truth of two clusters. The dashed ellipses represent clustering results. The clustering result in (a) may be far from satisfactory, since the points are clustered without click session information. The points in a solid-line ellipse in (b) compose a session. In (c), the points in the same sessions are connected mutually. In (d), the bigger point represents the goal image re-sampled from the points in a session. With the help of click session information, the points in (c) and (d) can be clustered more appropriately.

space when clustering. For example, the shapes of the clusters for green apples, red apples, and red laptops are spherical as shown in Fig. 7. The edge connecting two points means that these two images appear simultaneously in at least one session (i.e., some past users thought that these two images should be in one cluster). Therefore, the clusters for green apples and red apples will be merged under the guidance of users and the shape of the new cluster green and red apples (i.e., one of user search goals) will turn into strip. Therefore, the cluster shape of a user search goal can be arbitrary. Normally, the traditional methods, such as k -means clustering and affinity propagation (AP) clustering [26] are improper to handle these arbitrary-shape situations. However, with the introduction of spectral clustering, these situations can be suitably addressed.

4) *Step 4*: A CR-based approach is used to optimize the number of user search goals. When clustering, we first set the number of clusters k to be several probable values. Then we evaluate the clustering performances for each value of k according to the CR-based evaluation criterion. Finally, we choose the optimal value of k to be the number of user search goals. The detailed processes of Step 3 and Step 4 will be described in Section VI.

IV. EDGE-RECONSTRUCTION-BASED STRATEGY TO COMBINE IMAGE VISUAL INFORMATION WITH CLICK SESSION INFORMATION

We model the clicked images with similarity graph [25]. The vertices are the images and each edge is the similarity between two images. In the edge-reconstruction-based strategy, both image visual information and click session information are

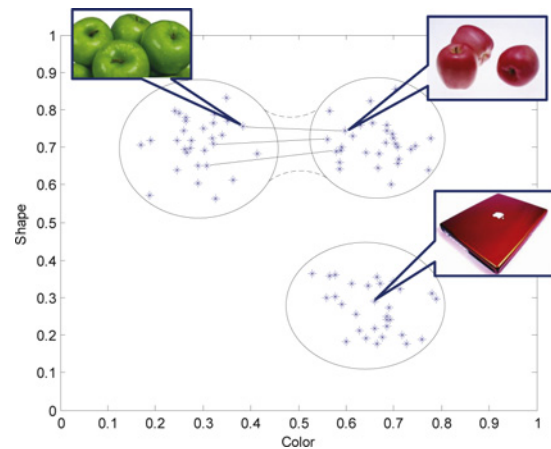


Fig. 7. Image clustering under the guidance of users. Each point represents an image. The x -axis and y -axis represent color and shape features of the image, respectively. The shapes of the clusters for green apple, red apple and red laptop are spherical. The edge connecting two points means that these two images appear simultaneously in at least one session.

utilized to compute similarities. Click session information can serve as a kind of semi-supervised information [27] for more precise clustering. In this section, we will first introduce the notion of the similarity graph. Then, we will describe the two steps for establishing the similarity graph. That is, Step 1, establishing the initial graph using the visual information of the clicked images, and Step 2, reconstructing the edges with click session information as semi-supervised information.

A. Notion of the Similarity Graph

The similarity graph $G = (V, E)$ is an undirected graph with vertex set $V = (v_1, \dots, v_n)$. Each edge between two vertices v_i and v_j carries a non-negative weight $w_{ij} \geq 0$. The weighted adjacency matrix of the graph is the matrix $\mathbf{W} = (w_{ij})_{i,j=1,\dots,n}$. If $w_{ij} = 0$, this means that the vertices v_i and v_j are not connected by an edge. As G is undirected, we require $w_{ij} = w_{ji}$. The degree d_i of a vertex $v_i \in V$ is defined as (2). The degree matrix \mathbf{D} is defined as the diagonal matrix with the degrees d_1, \dots, d_n on the diagonal

$$d_i = \sum_{j=1}^n w_{ij}. \quad (2)$$

B. Establishing the Initial Similarity Graph Using Image Visual Information

In the first step, we establish the initial similarity graph using the visual information of the clicked images. Let each clicked image for a query in user click-through logs be a vertex in V . Then the weight of the edge between two vertices v_i and v_j is the similarity between these two images s_{ij} as follows:

$$w_{ij} = s_{ij}. \quad (3)$$

The similarity between two images can be computed as the cosine score of their feature vectors as

$$s_{ij} = \cos(I_i, I_j) = \frac{I_i \cdot I_j}{|I_i| |I_j|} \quad (4)$$

where I_i is the normalized feature vector of the i th image.

C. Reconstructing the Edges With Click Session Information

As mentioned, since the existing image visual features may not precisely represent image semantics, the initial graph from the first step may still be less proper in inferring user search goals. Thus, after establishing the initial similarity graph, we further propose to utilize click session information as semi-supervised information for reconstructing the edges in the initial similarity graph. Although we do not know which cluster the clicked images in a session should be categorized into, the past users tell us that these clicked images should be in one cluster. Therefore, we connect the clicked images in a session as shown in Fig. 6(c) and propose another similarity metric between the images by utilizing click session information as follows:

$$s'_{ij} = \begin{cases} \frac{u_{ij}}{\beta} & u_{ij} < \beta \\ 1 & u_{ij} \geq \beta, \end{cases} \quad (5)$$

where u_{ij} is the number of the users who clicked the images v_i and v_j simultaneously. The constant β is for normalization. In this paper, we set β to be 10th number of all the users for a query. That is to say, if more than 10% users clicked the images v_i and v_j simultaneously, we consider that these two images are very similar and the similarity between these two images is set to be 1.

Then we update the similarity graph by adding s'_{ij} into (3) as

$$w_{ij} = \alpha s_{ij} + (1 - \alpha) s'_{ij} \quad (6)$$

where α is the coefficient to adjust the importance of two kinds of similarity metrics. For extreme, when α is 0, the similarity graph is totally determined by click session information and the feature representation of the vertices is no longer needed. This brings big flexibility to our approach.

V. GOAL-IMAGE-BASED STRATEGY TO COMBINE IMAGE VISUAL INFORMATION WITH CLICK SESSION INFORMATION

In the previous section, we have described the edge-reconstruction-based strategy that utilizes click session information to reconstruct the edges in the similarity graph. In this section, we propose another strategy, namely, goal-image-based strategy, which utilizes click session information to reconstruct the vertices in the similarity graph. In the following, we will first introduce the notion of goal image to explain why we re-sample the clicked images. Then, we will present how to re-sample a session (i.e., the clicked images in a session) into a goal image.

A. Goal Image

In image search, when users submit a query, they will usually have some vague figures or concepts in their minds as shown in Fig. 8. For the query ‘‘apple,’’ some users want to search the fruit apple. They usually know what an apple looks like. The shape should be round and the color should be red or green, etc. These are the common attributes (i.e., visual patterns) of the fruit apple to distinguish the fruit



Fig. 8. Goal images. For the query apple, users will have different vague figures or concepts in their minds. We name these vague figures goal images, which visually reflect users’ information needs.

apple from other things. Meanwhile, other users may want to search the computer or the cell phone of Apple Inc. These two search goals also have their own visual patterns. Therefore, users will use these vague figures consisting of those particular visual patterns in their minds rather than external texts to decide whether an image satisfies their needs. We name these vague figures goal images. They can visually reflect users’ information needs in image search. If we can obtain the goal image of each user, it will be feasible to infer user image-search goals by simply clustering all users’ goal images. However, the goal images in user minds are latent and not expressed explicitly. Therefore, we propose to re-sample the clicked images in a session as a whole (i.e., the session) into the goal image. Note that since each goal image represents a user, another advantage of re-sampling sessions into goal images is that we can easily obtain the distributions of the users having different search goals. More specifically, the search goal distribution of one query can be calculated by the ratio of the goal image number in one cluster against the number of all the goal images. With this distribution, we can know what search goals are more frequently searched for a specific query.

B. Re-Sampling Sessions Into Goal Images

Although goal images in user minds are latent, we can approximate them by mining user click-through logs since the clicked images in a session represent what a user needs. A session in image search is a series of clicked images for a query to satisfy a single information need of the user as shown in Fig. 2. Therefore, we re-sample the clicked images in a session by combining them into a super-image (i.e., goal image) as shown in Fig. 6(d).

There are two strategies to combine the images: feature fusion and image fusion. In Fig. 9, feature fusion (i.e., late fusion) first extracts features from each image in the session and then averages the features to generate \mathbf{f}_F . The disadvantage of feature fusion is that there is information loss when averaging. Therefore, we also propose image fusion (i.e., early fusion) that first collages the images and then extracts the feature of the collage \mathbf{f}_I . However, image fusion is only suitable for global features since global features reflect the overall statistics of the image collage and will not be affected by the image order in the collage. Comparatively, since local features mainly depend on the local statistics and locations in the image, the image order in the image collage will affect the fusion result if image fusion is used. Thus, most local features are only

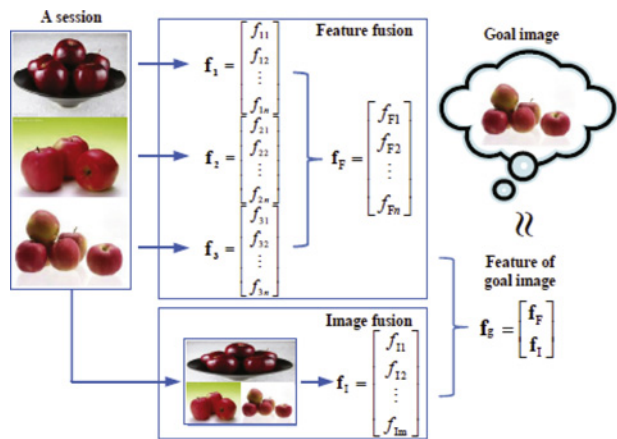


Fig. 9. Re-sampling the clicked images in a session into the goal image by two fusion strategies.

suitable for feature fusion. Therefore, in our paper, we fuse the local features (i.e., CMG and LBP) by feature fusion and fuse the global features (i.e., CC, CT and EAC) by image fusion. Finally, we concatenate f_F and f_I to get the feature representation of the goal image f_g .

There are several methods to create a picture collage from a group of images [28]. In this paper, we simply put the images into one frame to get the mosaic collage. Note that since the edge pixels between two images in a mosaic collage may have sudden changes, we do not include the features extracted from these edge pixels.

VI. CLUSTERING THE CLICKED IMAGES UNDER THE IMPLICIT GUIDANCE OF USERS WITH SPECTRAL CLUSTERING

The clicked images combined with click session information have been modeled as the similarity graph. In the edge-reconstruction-based strategy, the weighted adjacency matrix \mathbf{W} is obtained from (6). In the goal-image-based strategy, the vertices become the goal images and the edge weight is the cosine similarity between two goal images. Thus, a clustering process can be applied on the image graph to infer user search goals. In this section, we propose to introduce spectral clustering [25] to perform clustering. Furthermore, since the number of the clusters for queries may vary, it is also important to develop new methods to deal with this varying number of cluster problem. Therefore, we further propose a classification risk based approach to optimize the number of the clusters.

A. Spectral Clustering on the Similarity Graph

Basically, spectral clustering finds a partition of the similarity graph such that the edges between different groups have very low weights and the edges within a group have high weights. The spectral clustering algorithm used in this paper is described in Algorithm 1, where ε is a threshold to get the ε -neighborhood graph [25] and \mathbf{D} is the degree matrix as defined in Section V-A. In this paper, we set ε to be the average value of the weights $(w_{ij})_{i,j=1,\dots,n}$.

Algorithm 1 Spectral clustering

Require: The weighted adjacency matrix $\mathbf{W} = (w_{ij})_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

- 1: Let $w_{ij} = 0$ when $w_{ij} < \varepsilon$.
- 2: Compute the unnormalized Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$.
- 3: Compute the first k eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ of \mathbf{L} .
- 4: Let $\mathbf{U} \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ as columns.
- 5: For $i = 1, \dots, n$, let $\mathbf{y}_i \in \mathbb{R}^k$ be the vector corresponding to the i th row of \mathbf{U} .
- 6: Cluster the points $(\mathbf{y}_i)_{i=1,\dots,n}$ in \mathbb{R}^k with the k -means algorithm into clusters C_1, \dots, C_k .

Ensure: Clusters A_1, \dots, A_k with $A_i = \{j | y_j \in C_i\}$.

We propose to introduce spectral clustering in our case due to the following three reasons.

- 1) Spectral clustering can adaptively fit arbitrary cluster shapes. Therefore, it can work well in our case where the cluster shapes for different user search goals are varying, as described in Section III, Step 3.
- 2) In the edge-reconstruction-based strategy, since the similarity values are not only decided by the visual feature of the images as in (4), but are also decided by the click session information as in (6), it is difficult for k -means-like clustering methods to perform proper clustering with this kind of similarity information. However, since spectral clustering is able to work on graphs as long as the edge weights are available without considering how these weights are calculated, it is suitable to cluster the similarity graph in our case.
- 3) Furthermore, spectral clustering also has the advantage of enabling the clustering based on only click session information. For example, in the extreme case when the visual information of the vertex (i.e., image) is totally unreliable and the edge weight in the similarity graph is totally decided by the click session information (i.e., α in (6) is 0), the traditional k -means clustering will fail to work while the spectral clustering can still work properly.

Furthermore, note that in order to make the inference of user search goals robust, the isolated points (i.e., the vertices with very low degrees) are excluded before clustering. In the clustering process, since we do not know the exact number k of user search goals for each query, we propose to set k to be five different values (1, 2, \dots , 5) and perform clustering based on these five values, respectively. After clustering, we use those images with the highest degrees in each cluster to represent one of user image-search goals as shown in Fig. 10. The way to determine the best value of k will be described in the next subsection.

B. Classification Risk-Based Approach to Determine the Number of User Search Goals

We first develop a click-classification incoherence (CCI) metric to implicitly evaluate the performance of clustering (i.e., the performance of user-goal inference) by utilizing click



Fig. 10. Images with the highest degrees in the clusters to represent user image-search goals for the query “apple.”

session information. For a session after denoising, we classify each image in the session into one of k classes according to the clustering result. If clustering is proper, all the images in a session should be categorized into one class since the images clicked by one user should reflect the same search goal. If it is not the case, user clicks and our classification are incoherent. This incoherence can be calculated by

$$CCI = \frac{1}{T} \sum_{i=1}^T \frac{S_i - L_i}{S_i} \quad (7)$$

where T is the total number of the users submitting the same query (i.e., the sessions for the same query), S_i is the number of the images in the i th session and L_i is the number of the images in the largest class. The largest class is the class that has the most images when classifying the images in a session. For example, we use the denoised session Fig. 4(b) to compute $\frac{S_i - L_i}{S_i}$. That user considered that these six images should be in one cluster. If we categorize four of them into Class A and the other two into Class B, Class A is the largest class and for this session $\frac{S_i - L_i}{S_i} = \frac{6-4}{6} = 0.333$. In this situation the incoherence value between user clicks and our classification is 0.333.

Smaller k usually reduces the incoherence. In an extreme case, when we simply set $k = 1$ for all the queries, CCI will always be 0. For those non-ambiguous queries, if we categorize all the images into one class, the difference among the images in this class may be small. In this situation, $k = 1$ is reasonable. However, for those ambiguous queries, if we still categorize all the images into one class, the difference among the images in this class becomes very large. In this situation, $k = 1$ is not reasonable. Therefore, we further introduce the intra-class distance (ID) based on the adjacency matrix to revise the evaluation criterion. We classify all the clicked images into k clusters according to the clustering result. If clustering is proper, the image pairs in the same cluster should have short distances. We define ID as the average distance of all the image pairs in each cluster as

$$ID = \frac{1}{\sum_{m=1}^k \frac{N_m(N_m-1)}{2}} \sum_{m=1}^k \sum_{i,j=1,i < j}^{N_m} (1 - w_{ij}^{(m)}) \quad (8)$$

where N_m is the number of the clicked images in the m th cluster, $\sum_{m=1}^k \frac{N_m(N_m-1)}{2}$ is the total number of the image pairs in k clusters, and $w_{ij}^{(m)}$ is the edge weight of one image pair in the m th cluster. Bigger k usually reduces the intraclass distance.

Finally, CR , which represents the risk by improperly classifying the images according to the inferred user goals, consists of CCI and ID as follows:

$$CR = \lambda \cdot CCI + (1 - \lambda) \cdot ID \quad \lambda \in [0, 1]. \quad (9)$$

Thus, we cannot always set the number of user search goals to be 1 since ID could be very large. We choose the value of k when CR is the smallest. In order to determine the value of λ , we select 20 queries and empirically decide the number of user search goals of these queries. Then, we cluster the images and compute CR for different cluster numbers. We tune the value of λ to make CR the smallest when letting the number of clusters accord with what we expected for most queries. At last, we set λ to be 0.2 in this paper.

VII. EXPERIMENTS

In this section, we will show the experimental results of our proposed method. The dataset that we used is the query logs from one of the most popular commercial image search engine [19]. We randomly selected 100 queries that have more than 1000 clicks during one day. There are average 650 unique clicked images for a query. The average number of sessions for a query and the average number of clicks for a session is 100 and 20, respectively. Note that in practice, most queries have more than 100 sessions (i.e., a query searched and clicked by 100 times). Even if there are less than 100 sessions in one day, the data collection duration can be enlarged to collect enough sessions.

All the clicked images were downloaded according to the image URLs in query logs. In order to further compare our method with text based method, we additionally collect the external texts of the images for these 100 queries. In the following, we will first introduce the methods being compared. Then, we will give the quantitative comparisons among different methods. Finally, the illustrative examples and a subjective user evaluation experiment will be given.

A. Methods Being Compared

We compare the following five non-text methods to demonstrate the effectiveness of combining image visual information and click session information for inferring user image-search goals.

- 1) V_I_K (Visual_Image_Kmeans): clustering the clicked images with image visual information and k -means clustering.
- 2) V_I_S (Visual_Image_Spectral): clustering the clicked images with image visual information and spectral clustering.
- 3) C_I_S (Click_Image_Spectral): clustering the clicked images with click session information and spectral clustering.
- 4) VC_G_S (Visual-Click_Goal-image_Spectral): clustering the goal images, which are obtained by resampling the sessions with both image visual information and click session information, with spectral clustering.
- 5) VC_I_S (Visual-Click_Image_Spectral): clustering the clicked images by using both image visual information and click session information (as semi-supervised information) with spectral clustering.

Furthermore, we also compare our method with the following three text based methods (i.e., using the images' external

textual information for inferring user image-search goals).

- 1) T_Zha (Text_Zha): selecting the keywords (i.e., the terms representing user goals) by satisfying relatedness and informativeness according to Zha's work [4], and using the images whose external texts containing the keywords to visually represent user goals.
- 2) CT_I_S (Click-Text_Image_Spectral): clustering the clicked images by using the external texts (i.e., modeling the images with vector space model and computing the similarities with cosine similarity) and click session information with spectral clustering. Note that this method is similar to our VC_I_S algorithm and the only difference is that the textual features are used to take the place of the visual features.
- 3) VCT_I_S (Visual-Click-Text_Image_Spectral): clustering the clicked images by using the external texts, the image visual information, and the click session information with spectral clustering. Note that this is the extension of our VC_I_S algorithm by including the textual information (i.e., combining all the textual, visual, and click-through information for inferring user goals).

B. Quantitative Comparisons

Since the classification risk based evaluation criterion can effectively reflect user goal inference performance when deciding the number of user goals, we will use this criterion to quantitatively evaluate the performance of different user goal inference methods.

We first compare the five non-text based methods over all the 100 queries. In Fig. 11, each point represents *CCI* and Intradistance (*ID*) of a query. If user image-search goals are inferred properly, *CCI* and *ID* should be both small and the point should tend to be at the lower left corner. Fig. 11(a) compares V_I_S with V_I_K. We can see that the points of V_I_S are closer to the lower left corner comparatively. Therefore, V_I_S is better than V_I_K, which demonstrates that spectral clustering is more proper than *k*-means clustering when clustering the clicked images to infer user image-search goals. Fig. 11(b) and (c) compares VC_I_S with V_I_S and C_I_S, respectively. We can see that VC_I_S is better than either of the two methods, which demonstrates that clustering the clicked images with both image visual information and click session information is better than using only one of them. Fig. 11(d) compares the two strategies of our method and Fig. 11(e) compares V_I_S with C_I_S.

Then, we compare the average *CCI*, *ID*, and *CR* quantitatively for all the 100 queries among the five non-text methods as shown in Table 1. We can see that the average *CR*s of our two methods are the lowest. The *CR* of VC_I_S is the lowest, 14.29%, 8.86%, 14.03%, and 2.37% lower than the ones of V_I_K, V_I_S, C_I_S, and VC_G_S, respectively.

Comparing V_I_S and C_I_S, we can have the following observations.

- 1) Only using click session information can still achieve comparable performance as using image visual information. This implies the powerfulness of click session information in inferring user search goals.

TABLE I
AVERAGE *CR*, *CCI*, AND *ID* COMPARISONS
AMONG FIVE NON TEXT METHODS

Method	Average <i>CR</i>	Average <i>CCI</i>	Average <i>ID</i>
V_I_K	0.336	0.106	0.393
V_I_S	0.316	0.092	0.372
C_I_S	0.335	0.081	0.398
VC_G_S	0.295	0.087	0.347
VC_I_S	0.288	0.084	0.339

- 2) The average *CR* of C_I_S is comparatively higher than the one of V_I_S. This is because click session information in our dataset is still not extremely large. For example, in our dataset, there are average 650 unique clicked images for a query (i.e., 650 vertices in the similarity graph). Since the average number of sessions for a query and the average number of clicks for a session is 100 and 20 respectively in our dataset, click session information only creates about 9% edges in the similarity graph (note that a big large portion of these edges may even overlap). It is expected that the *CR* of C_I_S will be further decreased with more click-through information.
- 3) Since the effectiveness of click session information is proportional to the scale of user click data, it is expected that the performance of C_I_S as well as our VC_I_S/VC_G_S methods can be further improved with more user click data. Comparatively, the improvement by image visual information with the increased visual data may become less obvious.

We summarize the comparison between our VC_I_S and VC_G_S methods in the following.

- 1) The proposed edge-reconstruction-based strategy has the advantage of performing user goal inference with only click session information (i.e., by setting α to be 0 in (6)). By this way, a reliable result can still be expected even when image visual information is inaccessible or unreliable. Comparatively, both click session information and image visual information are required for the goal-image-based strategy.
- 2) The goal-image-based strategy has the advantage of achieving the useful distribution statistics of different search goals for a query while such information is inconvenient to obtain for the edge-reconstruction-based strategy.
- 3) From Table I, the user goal inference performance of the edge-reconstruction-based strategy is slightly better than the goal-image-based strategy in *CR* comparison. However, as mentioned, the performance of the edge-reconstruction-based strategy may be further improved if we can collect more clicks to update the similarity graph.

Furthermore, we also compare our method (VC_I_S) with three text based methods over all the 100 queries as shown in Table II. We have the following observations.

- 1) Comparing T_Zha and CT_I_S, we can see that click session information is also helpful when using the textual information to infer user image search goals.

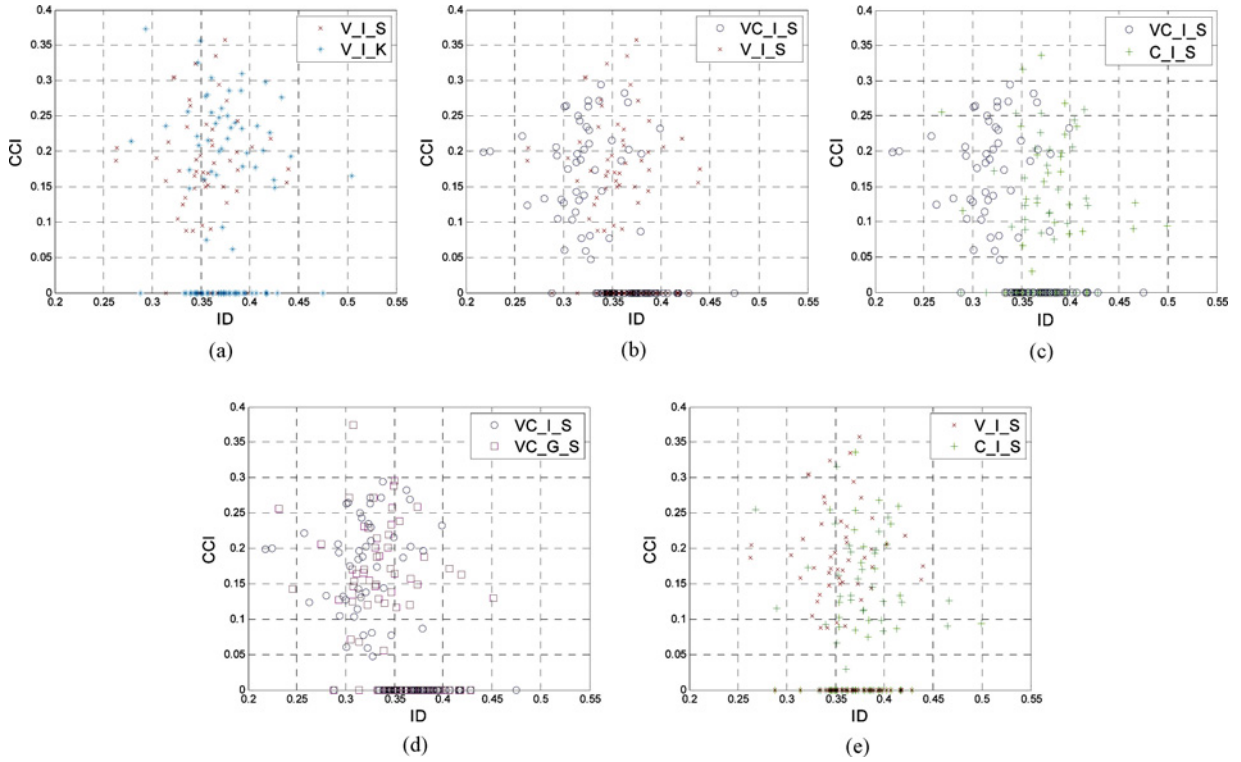


Fig. 11. *CCI* and *ID* comparisons among five nontext methods. The *x*-axis of the points represent *ID* and the *y*-axis represents *CCI*.

- 2) Comparing VC_I_S with T_Zha and CT_I_S, we can see that our VC_I_S method works better than the text based methods. This implies that by including the user click-through information, the semantic gap between the visual information and image semantics can be greatly narrowed, making the visual information more powerful than the external textual information in representing image-search goals. Besides, the visual information also has the advantage of differentiating search goals with different visual patterns. This point will be described in detail in the next subsection. Moreover, note that the external texts used in our experiments (from Baidu image [19]) were usually short and could not tell the difference between different goals in detail. Comparatively, since the tags in Zha’s work (from Flickr [18]) were selected to describe the images, their tags were more precise and could depict the images more accurately. This is another reason why the average *CR*s of T_Zha and CT_I_S are higher than the ones of V_I_S and VC_I_S respectively in our experiments.
- 3) VCT_I_S usually has the lowest average *CR*. It shows that when textual information is available and reliable, including the textual information into our algorithm (i.e., combining textual, visual, and click information) can further improve the performance. The extended version of our algorithm by combining textual, visual, and click information has the best performance.

C. Illustrative Examples and Subjective User Evaluation

In this subsection, we will first show the illustrative examples for different user goal inference methods. For the

TABLE II
AVERAGE *CR*, *CCI*, AND *ID* COMPARISONS AMONG OUR METHOD AND THREE TEXT BASED METHODS

Method	Average <i>CR</i>	Average <i>CCI</i>	Average <i>ID</i>
T_Zha	0.329	0.096	0.387
CT_I_S	0.313	0.088	0.369
VC_I_S	0.288	0.084	0.339
VCT_I_S	0.277	0.078	0.327

edge-reconstruction-based strategy (i.e., VC_I_S), we choose the clicked images with the highest degree (i.e., d_i in (2)) in each cluster to represent one of user image-search goals. Fig. 12 shows part of the results. While for the goal-image-based strategy (i.e., VC_G_S), the center point of the cluster can be represented by a feature vector. We choose the clicked image closest to the cluster center as the image example to represent the user image-search goal as shown in Fig. 13. The distributions of different search goals are also given beside the image examples. From Figs. 12 and 13, we can see that both of the two strategies of our method can infer user image-search goals properly and these two strategies can achieve similar clustering results.

Table III shows the query statistics over different cluster (goal) numbers and different query types. Each cell is the number of queries. As mentioned by Fig. 1, we can see from Table III that most queries have multiple search goals and different types of reasons can make the query contain multigoals.

Furthermore, we show some illustrative examples for comparing non-text based methods and text based methods in Fig. 14. Fig. 14 further shows the following.

query	Different user image-search goals	query	Different user image-search goals
1. apple		9. Liu Xiang	
2. bird's nest		10. basketball	
3. Millet		11. soccer	
4. Bumblebee		12. lotus	
5. coffee		13. water-drop	
6. tea		14. lantern	
7. diamond		15. leaf	
8. watermelon		16. fish	

Fig. 12. Inferred user image-search goals by the edge-reconstruction-based strategy of our method (i.e., VC-I-S).

query	Different user image-search goals and their distributions	query	Different user image-search goals and their distributions
1. apple		4. coffee	
2. bird's nest		5. soccer	
3. Bumblebee		6. lotus	

Fig. 13. Inferred user image-search goals and their distributions by the goal-image-based strategy of our method (i.e., VC-G-S).

TABLE III
QUERY STATISTICS OVER DIFFERENT CLUSTER NUMBERS AND DIFFERENT QUERY TYPES

Numbe of clusters (goals)					
Query types	1	2	3	4	5
Single goal query	37	0	0	0	0
Multi concepts	0	5	4	2	0
Multi forms	0	11	6	2	1
Multi representations	0	15	11	5	1

- 1) Only using single type of information (i.e., V-I-S, C-I-S, or T-Zha) cannot achieve satisfying search goals. V-I-S wrongly clustered the images according to brightness for the query bird's nest. C-I-S omitted the iPhone goal for the query apple due to fewer clicks on iPhone images. T-Zha obtained redundant clusters for the query bird's nest.
- 2) Comparatively, by combining click, visual and textual information, (i.e., VC-I-S and VCT-I-S), the user goals can be inferred properly for both queries.

Finally, we perform a user evaluation experiment to assess whether the inferred search goals are coherent with user

method	Different user image-search goals for the query "apple"	Different user image-search goals for the query "bird's nest"
V-I-S		
C-I-S		
T-Zha	fruit logo iphone 	stadium egg tree
VC-I-S		
VCT-I-S		

Fig. 14. Illustrative examples for comparing non-text based methods and text based methods.

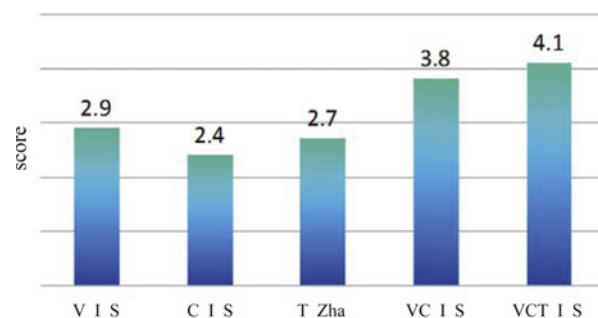


Fig. 15. User judgements for five methods.

judgements. We invited 30 users evaluate the user goal inference results from the five methods, including 15 undergraduates, eight graduate students, four Ph.D. candidates, and three staff members. Their ages ranged from 20 to 52. Twenty two of them were males and eight were females. They covered a wide variety of majors, such as computer science, maths, economics, literature, and so on. Twenty four of them had the experience of using image search engines and six of them were unfamiliar with image search. We let each user grade the performances of five goal inference methods for ten queries (i.e., each of the 100 queries was evaluated by three different users). The results from different methods were randomly ordered such that "which result belongs to which method" was unknown to users. Five scores were provided: 1 for not satisfied, 2 for slightly satisfied, 3 for neutral, 4 for satisfied, and 5 for very satisfied. The scores were averaged over all queries and over all users for each method. Fig. 15 shows the results. From Fig. 15, we can see that our method (i.e., VC-I-S) can get satisfying results to the users. And by combining the textual information, the extension of our method (i.e., VCT-I-S) can achieve the most satisfying results.

VIII. CONCLUSION

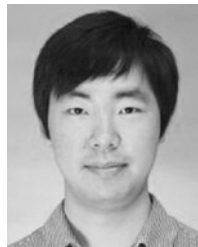
In this paper, we proposed to leverage click session information and combine it with image visual information to infer user image-search goals. Click session information can serve as the implicit guidance of the past users to help clustering. Based on

this framework, we proposed two strategies to combine image visual information with click session information. Furthermore, a click-classification incoherence based approach was also proposed to automatically select the optimal search goal numbers. Experimental results demonstrated that our method can infer user image-search goals precisely.

It is worth noting that the proposed method in this paper focused on analyzing a particular query appearing in the query logs. Inferring user image-search goals for those popular queries can be very useful, and our proposed method can also be extended for a new query (not appearing in the query logs). For example, we can infer user image-search goals for a group of similar queries instead of a particular query. The new query will be classified into a query group at first. Then the user search goals for the query group can be considered as the ones for this new query.

REFERENCES

- [1] D. Tjondronegoro, A. Spink, and B. Jansen, *A Study and Comparison of Multimedia Web Searching: 1997–2006*, vol. 60, no. 9. Wiley Online Library, 2009, pp. 1756–1768.
- [2] B. Jansen, A. Spink, and J. Pedersen, “The effect of specialized multimedia collections on web searching,” *J. Web Eng.*, vol. 3, no. 3–4, pp. 182–199, 2004.
- [3] B. Jansen, A. Goodrum, and A. Spink, “Searching for multimedia: Analysis of audio, video and image web queries,” *World Wide Web*, vol. 3, no. 4, pp. 249–254, 2000.
- [4] Z. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T. Chua, and X. Hua, “Visual query suggestion: Toward capturing user intent in internet image search,” *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 6, no. 3, p. 13, 2010.
- [5] H. Chen and S. Dumais, “Bringing order to the web: Automatically categorizing search results,” in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2000, pp. 145–152.
- [6] S. Wan, Y. Xue, X. Yu, F. Guan, Y. Liu, and X. Cheng, *ICTNET at Web Track 2011 Diversity Task*. MD, USA: National Instit. Standards Technology, 2011.
- [7] R. Santos, C. Macdonald, and I. Ounis, “Exploiting query reformulations for web search result diversification,” in *Proc. 19th Int. Conf. World Wide Web*, 2010, pp. 881–890.
- [8] U. Lee, Z. Liu, and J. Cho, “Automatic identification of user goals in web search,” in *Proc. 14th Int. Conf. World Wide Web*, 2005, pp. 391–400.
- [9] D. Rose and D. Levinson, “Understanding user goals in web search,” in *Proc. 13th Int. Conf. World Wide Web*, 2004, pp. 13–19.
- [10] X. Li, Y. Wang, and A. Acero, “Learning query intent from regularized click graphs,” in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, vol. 339, 2008, p. 346.
- [11] X. Wang and C. Zhai, “Learn from web search logs to organize search results,” in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2007, pp. 87–94.
- [12] Z. Lu, H. Zha, X. Yang, W. Lin, and Z. Zheng, “A new algorithm for inferring user search goals with feedback sessions,” 2011.
- [13] Z. Lu, X. Yang, W. Lin, X. Chen, and H. Zha, “Inferring users’ image-search goals with pseudo-images,” in *Proc. IEEE Visual Commun. Image Process.*, 2011, pp. 1–4.
- [14] D. Cai, X. He, Z. Li, W. Ma, and J. Wen, “Hierarchical clustering of www image search results using visual, textual and link information,” in *Proc. 12th Annu. ACM Int. Conf. Multimedia*, 2004, pp. 952–959.
- [15] P. Enser and C. Sandom, “Toward a comprehensive survey of the semantic gap in visual image retrieval,” in *Proc. Image Video Retrieval*, 2003, pp. 163–168.
- [16] R. Jones and K. Klinkner, “Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs,” in *Proc. 17th ACM Conf. Inform. Knowl. Manage.*, 2008, pp. 699–708.
- [17] D. Shen, J. Sun, Q. Yang, and Z. Chen, “Building bridges for web query classification,” in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 2006, pp. 131–138.
- [18] Flickr [Online]. Available: <http://www.flickr.com>.
- [19] Baidu Image Search [Online]. Available: <http://image.baidu.com>.
- [20] *Google Image Search* [Online]. Available: <http://images.google.com>.
- [21] J. Carbonell and J. Goldstein, “The use of MMR, Diversity-based reranking for reordering documents and producing summaries,” in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval*, 1998, pp. 335–336.
- [22] T. Arni, J. Tang, M. Sanderson, and P. Clough, “Creating a test collection to evaluate diversity in image retrieval,” in *Proc. Beyond Binary Relevance: Preferences, Diversity Set-Level Judgments*, 2008.
- [23] H. Cheng, K. Hua, and K. Vu, “Leveraging user query log: Toward improving image data clustering,” in *Proc. Int. Conf. Content-Based Image Video Retrieval*, 2008, pp. 27–36.
- [24] B. Poblete, B. Bustos, M. Mendoza, and J. Barrios, “Visual-semantic graphs: Using queries to reduce the semantic gap in web image retrieval,” in *Proc. 19th ACM Int. Conf. Inform. Knowl. Manage.*, 2010, pp. 1553–1556.
- [25] U. Von Luxburg, “A tutorial on spectral clustering,” *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [26] B. Frey and D. Dueck, “Clustering by passing messages between data points,” *Sci.*, vol. 315, no. 5814, pp. 972–976, 2007.
- [27] N. Grira, M. Crucianu, and N. Boujemaa, “Unsupervised and semi-supervised clustering: A brief survey,” *A Review of Machine Learning Techniques for Processing Multimedia Content*, Report of the MUSCLE European Network of Excellence (FP6), 2004.
- [28] J. Wang, L. Quan, J. Sun, X. Tang, and H. Shum, “Picture collage,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recogn.*, vol. 1, 2006, pp. 347–354.



Zheng Lu received the B.S. degree from Fudan University, Shanghai, China, in 2005, and the M.E. degree from Shanghai Jiao Tong University, Shanghai, in 2008, in electrical engineering. He is currently pursuing the Ph.D. degree at the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University.

His current research interests include information retrieval, machine learning, and video processing.



Xiaokang Yang (M’00–SM’04) received the B.S. degree from Xiamen University, Xiamen, China, in 1994, the M.S. degree from the Chinese Academy of Sciences, Shanghai, China, in 1997, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000.

He is currently a Professor and the Vice Dean of the School of Electronic Information and Electrical Engineering, and the Deputy Director of the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University. From August 2007 to July 2008, he visited the Institute for Computer Science, University of Freiburg, Freiburg, Germany, as an Alexander von Humboldt Research Fellow. From September 2000 to March 2002, he was a Research Fellow at the Center for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a Research Scientist with the Institute for Infocomm Research (I2R), Singapore. He has published over 150 refereed papers, and has filed 30 patents. His current research interests include visual signal processing and communication, media analysis and retrieval, and pattern recognition.



Weiyao Lin received the B.E. and M.E. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2005, respectively, and the Ph.D. degree from the University of Washington, Seattle, WA, USA, in 2010, all in electrical engineering.

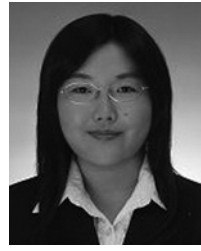
Since 2010, he has been an Assistant Professor with the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. His current research interests include video processing, machine learning, computer vision, video coding,

and compression.



Hongyuan Zha received the B.S. degree in mathematics from Fudan University, Shanghai, China, in 1984, and the Ph.D. degree in scientific computing from Stanford University, Stanford, CA, USA, in 1993.

He was a Faculty Member with the Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA, from 1992 to 2006, and from 1999 to 2001 was with Inktomi Corporation, CA, USA. His current research interests include computational mathematics, machine learning applications, and information retrieval.



Xiaolin Chen received the B.E. degree from the University of Electronic, Science and Technology of China, Chengdu, China, in 2004, and the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2012, all in electronic engineering.

Since 2012, she has been with the Nanjing Research Institute of Electronics Engineering, Nanjing, China. Her current research interests include image processing, computer vision, and pattern recognition.