Abstract — Inferring user search goals is very important in improving search engine relevance and user experience. Queries may not exactly represent the need of the user because of polysemy in keywords. At some times user may tend to form short queries so that the sense of search may be messed up on the whole. So far image is been searched as product indent most probably based on query classification. Initially they select some tag words as textual suggestions to satisfy relatedness and informativeness; however it depends on the precision of tags. So as to leverage the semantic gap between image features and image semantics a strategy to focus on click content logs is proposed. Both strategies click content information and click session information works drastically forward to sum up the gap between image visual features and textual content. The image results are then re-ranked in a diverse manner to provide a better idea about the semantic correlation among images and its number search goals of each query can be obtained by a classification risk approach.

Keywords — search goal, Classification Risk (CR), click content information, click session information

I. INTRODUCTION

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, the user image search goals as visual query suggestions to help users reformulate their queries during image search. Besides, categorize search results for image search according to the inferred user image-search goals to make it easier for users to browse. Furthermore, diversify and re-rank the results retrieved for a query 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 have been many researches for text search, few methods were proposed to infer user search goals in image search. 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 which contains that image and the surrounding texts in image search results and the tags given by users). 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 their 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 between the existing image features and the image semantics, inferring user image-search goals by visual information is still a big challenge. Therefore, additional information sources introduced to help narrowing these semantic gaps.

The contributions in this paper can be described as follows:

1) Propose a new framework that combines image visual information and click session information for inferring user image-search goals for a query. By this way, more precise image-search goals can be achieved.

2) Two strategies (i.e., the edge reconstruction- based strategy and the goal-image based strategy) proposed to effectively implement the process of combining image visual information with click session information. And also spectral clustering is introduced 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), further
A. Generate image visual information

Initially, the visual information is extracted 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.

B. Extract click session information

Then extract the click session information from user click-through logs. 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 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. If these noisy images are included, the click session information will become less meaningful.

C. Information combining

D. Spectral clustering

Spectral clustering algorithm is used to cluster the image graph which 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 space when clustering. For example, the shapes of the clusters “green apples”, “red apples” and “red laptops” are spherical. 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 “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 are improper to handle these arbitrary-shape situations. However, with the introduction of spectral clustering, these situations can be suitably addressed.

E. Classification risk based approach

A classification risk (CR) based approach is used to optimize the number of user search goals. When clustering, the number of clusters is set to k to be several probable values. Then evaluate the clustering performances for each value of k according to the CR based evaluation criterion. Finally, choose the optimal value of k to be the number of user search goals.

Image visual information is combined with click session information for further clustering by one of the two proposed strategies named as 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. 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.

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

The clicked images modeled with similarity graph. 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 utilized to compute similarities. Click session information can serve as a kind of semisupervised information for more
Then propose another similarity metric between the images by utilizing click session information as follows:

\[ w_{ij} = s_{ij} \]

(3)

where \( u_{ij} \) is the number of the users who clicked the images \( v_i \) and \( v_j \) simultaneously. The constant \( \beta \) is for normalization (normally to be ten percent of the 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, consider that these two images are very similar and the similarity between these two images is set to be 1.

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

The goal-image-based strategy which utilizes click session information to reconstruct the vertices in the similarity graph. Another strategy, namely

\[ \text{(2)} \]

where \( I_i \) is the normalized feature vector of the \( i \)-th image.

The similarity between two images can be computed as the cosine score of their feature vectors as Eqn. (2):

\[ s_{ij} = \cos(I_i, I_j) = \frac{I_i \cdot I_j}{||I_i|| ||I_j||} \]

In the first step, the initial similarity graph is established 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:

A. Establishing the initial similarity graph using image visual information

precisely clustering. Then, two steps are used 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.

B. Reconstructing the edges with click session information

\[ s_{ij} = \begin{cases} \frac{u_{ij} \cdot \beta}{\beta^2} & \text{if } u_{ij} < \beta^2 \\ 1 & \text{if } u_{ij} \geq \beta^2 \end{cases} \]

Fig. 2: Re-sampling the clicked images in a session into the goal image by two fusion strategies.
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. In the clustering process, since the exact number \( k \) of user search goals for each query is not known, propose to set \( k \) to be 5 different values (1, 2, ..., 5) and perform clustering based on these five values, respectively. After clustering, use those images with the highest degrees in each cluster to represent one of user image-search goals.

B. Classification risk based approach to determine the number of user search goals

Initially 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 session information. For a session after denoising, each image in the session is classified 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 classification are incoherent. This incoherence can be calculated by,

\[
CCI = \frac{1}{T} \sum_{i=1}^{T} \frac{S_i - L_i}{S_i},
\]

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.

Then introduce the Intra-class Distance (ID) based on the adjacency matrix to revise the evaluation criterion. All the clicked images are classified into \( k \) clusters according to the clustering result. If clustering is proper, the image pairs in the same cluster should have short distances. So ID is defined as the average distance of all the image pairs in each cluster as:

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

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 intra-class distance.

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

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

Thus, the number of user search goals cannot be set to 1 since \( ID \) could be very large. So, choose the value of \( k \) when \( CR \) is the smallest. In order to determine the value of \( \lambda \), 20 queries selected and empirically decide the number of user search goals of these queries. Then cluster the images and compute \( CR \) for different cluster numbers. The value of \( \lambda \) is tuned to make \( CR \) the smallest when letting the number of clusters accord with what is expected for most queries.

VI. EXPERIMENTAL EVALUATION

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

- **V_\text{I}_K** (Visual Image K-means): clustering the clicked images with image visual information and \( k \)-means clustering.
- **V_\text{I}_S** (Visual Image Spectral): clustering the clicked images with image visual information and spectral clustering.
- **C_\text{I}_S** (Click Image Spectral): clustering the clicked images with click session information and spectral clustering.
- **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.
- **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, the following three text based methods (i.e., using the images’ external textual information for inferring user image-search goals) is compared.
**T_Zha (Text Zha):** selecting the keywords (i.e., the terms representing user goals) by satisfying relatedness and informativeness, and using the images whose external texts containing the keywords to visually represent user goals.

**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 VC_I_S algorithm and the only difference is that the textual features are used to take the place of the visual features.

**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 VC_I_S algorithm by including the textual information (i.e., combining all the textual, visual, and click-through information for inferring user goals).

![Fig 3: User judgments' for five methods](image)

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. 3 shows the results. From Fig. 3, we can see that the method (i.e., VC_I_S) can get satisfying results to the users. And by combining the textual information, the extension of this method (i.e., VCT_I_S) can achieve the most satisfying results.

**VII. CONCLUSION**

To leverage click session information is proposed 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, two strategies proposed to combine image visual information with click session information. Furthermore, a click-classification incoherence based approach is also proposed to automatically select the optimal search goal numbers. Experimental results demonstrate that this method can infer user image-search goals precisely. It focuses on analyzing a particular query appearing in the query logs. Inferring user image-search goals for those popular queries can be very useful.

**REFERENCES**


