USING IMPLICIT RELEVANE FEEDBACK TO ADVANCE WEB IMAGE SEARCH*

En Cheng[1], Feng Jing[2], Mingjing Li[2], Weiying Ma[2], and Hai Jin[1]

[1]Huazhong University of Science & Technology, Wuhan, 430074, P.R. China
[2]Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China

ABSTRACT

Although relevance feedback has been extensively studied in content-based image retrieval in the academic area, no commercial web image search engine has employed the idea. There are several obstacles for Web image search engines in applying relevance feedback. To overcome these obstacles, we proposed an efficient implicit relevance feedback mechanism. The proposed mechanism shows advantage over traditional relevance feedback methods in the following three aspects. Firstly, instead of enforcing the users to make explicit judgment on the results, our method regards user's click-through data as implicit relevance feedback which release burden from users. Secondly, a hierarchical image search results clustering algorithm is proposed to semantically organize the search results. Using the clustering results as features, our relevance feedback scheme could catch and reflect users' search intention precisely. Lastly, unlike traditional relevance feedback user interface which hardily substitutes subsequent results for previous ones, our method employed friendly recommendation rather than substitution to let the user narrow down on the refined images. To evaluate the implicit relevance feedback mechanism, comprehensive user studies were performed.

1. INTRODUCTION

Image retrieval system is developing driven by the explosive growth of both World Wide Web and the number of digital images. As a result, there are now a number of Web image search engines, like Google [1], Picsearch [7], and Yahoo! [9], available for locating digital images. These images are automatically indexed by textual features such as their captions and surrounding texts rather than visual features. In fact, making effective use of textual features can render image retrieval by high-level concepts more efficient, and leverage mature techniques from text retrieval. The issues related to the design and implementation of a Web image search engine, such as data gathering and digestion, indexing, query specification, retrieval and similarity, Web coverage, and performance evaluation are thoroughly discussed in [5]. A common limitation of the existing Web image search engines is that their search process is passive, i.e. disregarding the informative interactions required for reaching good results between user and search engine. Therefore, there is an urgent need of an effective relevance feedback mechanism applied to image retrieval from the World Wide Web.

Relevance feedback, originally developed for information retrieval [3] is an online learning technique used to improve the effectiveness of information retrieval systems. Since its introduction into image retrieval in middle 1990's, it has attracted tremendous attention in the Content-Based Image Retrieval (CBIR) community and has been shown to provide dramatic performance improvement [8]. The main idea of relevance feedback is to let the user into the loop and guide the system. During retrieval process, the user interacts with the system and rates the relevance of the retrieved images, according to his/her subjective judgment. With this additional information, the system dynamically learns the user's intention, and gradually presents better results.

There are at least three key issues in Web image search engines when utilizing relevance feedback scheme. The three key issues are listed below:

- Real-time Requirement

Since the user is interacting with the search engine in real time, the relevance feedback mechanism should be sufficiently fast, and if possible avoid heavy computations over millions of retrieved images. Traditional relevance feedback methods, such as the query point movement method and the re-weighting method, are impractical for preserving interactivity.

- User convention

A straightforward way to get the user into the loop is to ask him/her to provide explicit feedback regarding the (ir)relevance of the current retrieval results. However, such process adds too much burden on a common user. A more

* This work was performed at Microsoft Research Asia.

feasible form of interaction is to distill information from user's implicit relevance feedback, e.g. click-through data.

- User interface

A friendly search UI should both provide the user a way to communicate with the retrieval system and enable the user view the search results smoothly and efficiently. There are two limitations of the traditional relevance feedback UI in Web application. One is that keeping the user waiting for a rather long period while refreshing the whole resulting page is really a big challenge to the user's patience. The other more severe limitation is that the relevance of the refined results could not be guaranteed. If a long period of waiting end up with unsatisfied results, the user will lose their confidence on the system and be frustrated.

In this paper, we proposed an efficient and effective method to address the above issues. Our method regards user's click-through data as implicit relevance feedback so that no extra burden will be put on the user. An efficient and effective image search results clustering algorithm is proposed to generate semantic features based on which the relevance feedback scheme works. To facilitate the relevance feedback algorithm, a new user interface is proposed. Instead of hardily substituting subsequent results for previous ones, representatives from refined queries are recommended to the user by side of the normally retrieved results.

The organization of the paper is as follows. Section 2 describes the proposed implicit relevance feedback scheme including both an image search result clustering algorithm and a new user interface. In Section 3, extensive user study results are presented and analyzed. Finally, we conclude in Section 4.

2. IMPLICIT RELEVANCE FEEDBACK MECHANISM

As the basis of the relevance feedback scheme, an efficient image search result clustering algorithm is proposed and will be discussed in section 2.1. Using user's click-through data as implicit feedback information, a new user interaction & interface is introduced in section 2.2.

2.1 Image Search Result Clustering (ISRC) algorithm

Learning candidate image cluster names

The candidate image cluster names are generated from two sources. One is the salient phrases extracted from the clustering results of Google's web page search [2]. The other is from the suggested phrases of an image search engine, i.e. Picsearch [7]. For the former, we use the algorithm proposed in [4]. [4] re-formalizes the clustering problem as a salient phrase ranking problem. Given a query and the ranked list of search results, it first parses the whole list of titles and snippets, extracts all possible phrases (n-grams) from the contents, and calculates several properties for each phrase such as phrase frequencies, document frequencies, phrase length, etc. A regression model learned from previous training data is then applied to combine these properties into a single salience score. The phrases are ranked according to the salience score, and the top-ranked phrases are taken as salient phrases. The salient phrases are further merged according to their corresponding documents. An online demo showing the algorithm of [4] is [6]. The resulting salient phrases are one source of the candidate image cluster names. On the other hand, Picsearch [7] will suggest up to 5 related phrases for each query. These suggested phrases are another source of the candidate image cluster names.

Merging and pruning cluster names

Given the candidate cluster names, a merging and pruning algorithm is utilized to obtain the final cluster names. Firstly, we merged the same or very similar candidates from different sources. Secondly, the synonyms of "images", e.g. "pictures" or "photos" are utilized to prune the candidate cluster names of possibly helpless clusters. Finally, the resulting candidate cluster names are used as queries to search an image search engine, e.g. Google image search [1] with the number of resulting images counted. The cluster names with too many or too few resulting images are further pruned. Each left cluster name corresponds to a cluster that contains the images returned by the search engine using the cluster name as query. The reduced thumbnails of top ranked images are used as representative images of the clusters. The proposed ISRC algorithm is shown in Fig. 1.

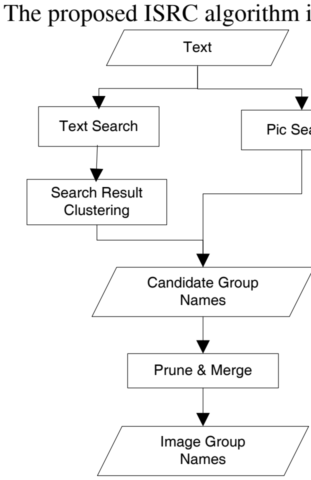

Fig. 1. Flowchart of image search result clustering algorithm.

2.2 User Interaction & Interface

When the user starts a search using a text query, the first-layer semantic groups are generated by utilizing the aforementioned clustering algorithm. Initially, without any interaction, the recommended images are selected from the top 6 semantic groups. Semantic groups are ranked according to the number of the contained images. More reasonable ranking criterion, such as ranking the groups according to the frequencies of the group names as queries could also be employed if the query log is available.

During typical relevance feedback process, the user's judgments on retrieved images are used to refine the query to get more relevant results. To preserve interactivity, the relevance feedback mechanism implemented in a search engine must operate in real time. In our system, if the user clicked one image for sake of his/her interest, the system will regard the click-through data as implicit relevance feedback and dynamically accomplishes the hierarchical clustering to give birth to a more precise range of semantic groups. The flowchart of our implicit relevance feedback mechanism is shown in Fig. 2. To make the best of interactions between user and search engine, a new user interface is introduced. A snapshot of the UI is shown in Fig. 3. The proposed interface consists of two view frames. The resulting images are listed in the right frame with 5 rows and 4 columns. The recommended images are shown in the left frame in a rolling mode. The recommended images are representatives from the second-layer semantic groups and the relevant images from the clicked semantic group. The more clicked images from the same semantic group, the more probably the refined semantic group meets the user's need. The most benefit from the interaction is the refined query especially when the user has no precise word to describe the specific characteristic in his mind.

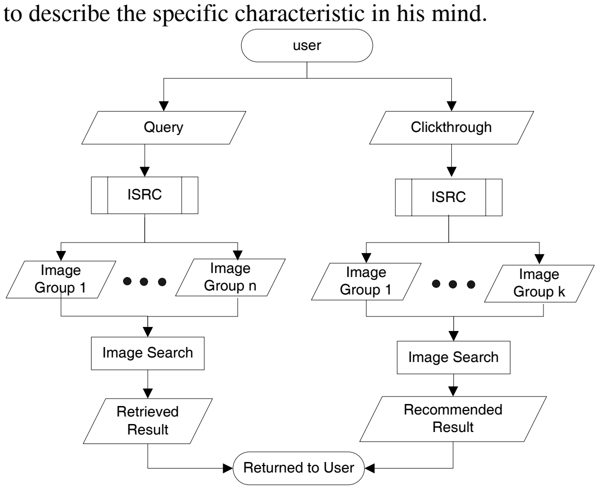

Fig. 2. Flowchart of implicit relevance feedback mechanism.

As we discussed earlier, efficiency issue is very critical for relevance feedback mechanism. The time cost of our ISRC algorithm is around 3s, so the user need wait around 3s for generating the first-layer semantic groups after inputting his/her query. With users' interaction, the second-layer semantic groups come into being during user's normally viewing period. Since the hierarchical refinement of queries for recommended images is independent from the displaying, uses are released from a long period of waiting for refreshing the whole resulting page. In this way, the real-time requirement could be satisfied.

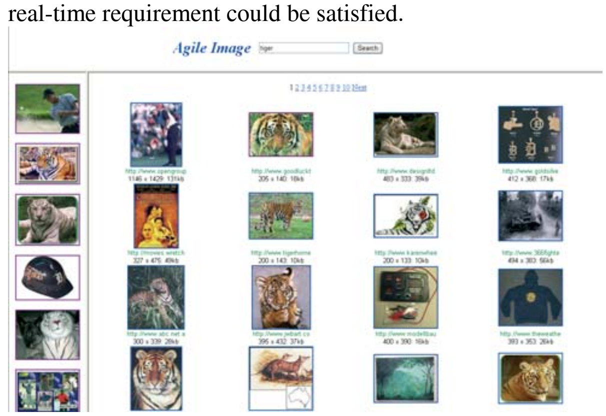

Fig. 3. Main user interface of Agile IMage (AIM).

3. USER STUDY

Participants & Apparatus

15 participants (7 female and 8 male) were asked to involve in all the following user study. The participants were all regular users of the Internet, searching for information very often. More than half of them search for images several times a week and others less than twice per week.

Each participant accomplished the test on a machine using Internet Explorer 6 of Windows XP with a 17 inch LCD monitor set at 1280*1024 pixels in 24-bit color. Data was recorded with multiple methods: 1) paper surveys after each task and at the end of user study; 2) behavioral logs (time stamps) and 3) server logs. One experienced usability analyst conducted the user study.

Tasks

To compare the proposed Agile IMage (AIM) system with existing image search engines, e.g. MSN image search [5], we defined three specific tasks. The first task is assuming you will make an introduction to your friend about traveling in Beijing, search for 10 representative images of Beijing. The second task is assuming you will design a booklet of Harry Potter and search for 10 material images. The third task is assuming you will design a homepage for NIKE products, search for 10 representative images. All 15 participants were involved in all three tasks.

Firstly, the average time used to complete the tasks was evaluated. The ANOVA test results of task 1-3 are shown in Table 1. From the results we can see that the search time of task 2 and 3 using AIM is obviously less than using MSN. When it comes to task 1 – searching for 10 representative images of Beijing, our participants have a clear vision of Beijing, so they can search the related places directly. As a result, the search time of task 1 has little difference between MSN and AIM.

After each task, participants completed a short questionnaire containing two questions. One is "Are you confident that the images you found are relevant images". The other is "Are you satisfied with the results". For the former, the participants are required to select from four options for MSN or AIM. The options are very confident, somewhat confident, unconfident and very unconfident. For the latter, similar options are used: very satisfied, somewhat satisfied, unsatisfied and very unsatisfied. To facilitate further statistical analysis, the results are quantitated to rating 4, 3, 2, and 1 respectively. The ANOVA tests were performed on both confidence value and satisfaction value. The results of task 1-3 are shown in Table 2 and Table 3. From the results we can see that both confidence and satisfaction values of AIM are higher than those of MSN.

Table 1. ANOVA test of search time of task 1-3.

Task	Engine	Mean	Variance	F(1,14)	P Value
1	MSN	185.9	785.4	7.36	0.017
	AIM	184.1	484.1		
2	MSN	209.7	800.6	6.75	0.025
	AIM	167.3	367.8		
3	MSN	180.4	906.7	6.47	0.038
	AIM	145.7	576.8		

Table 2. ANOVA test of confidence value of task 1-3.

Task	Engine	Mean	Variance	F(1,14)	P Value
1	MSN	3.7	0.32	14.78	0.004
	AIM	3.5	0.45		
2	MSN	2.8	0.62	23.34	0.036
	AIM	3.8	0.23		
3	MSN	3.0	0.85	17.84	0.002
	AIM	3.9	0.57		

Table 3. ANOVA test of satisfaction value of task 1-3.

Task	Engine	Mean	Variance	F(1,14)	P Value
1	MSN	2.3	1.23	10.54	0.001
	AIM	3.0	0.64		
2	MSN	3.3	0.87	19.75	0.004
	AIM	3.7	0.16		
3	MSN	3.1	0.64	14.73	0.002
	AIM	3.5	0.34		

Overall Comments

15 participants (7 female and 8 male) were asked to complete a questionnaire to provide overall comments on the user interface and the recommendation results.

Firstly, the participants were asked to state the cons of the interactive browsing. According to the comments, there are two limitations. The first limitation is that users need a guideline for the Agile Image in advance. The second limitation is that if the user has a clear description of what he wants, our interactive browsing will be inessential.

Secondly, the participants were asked to state the cons of the recommendation results. Most of the cons are about the ranking of the recommendation groups. As we discussed earlier, we rank the semantic groups according to the number of the contained images, which neglects the inherent inference knowledge of user's click-through. To make the recommendation more helpful, the inference knowledge should be utilized to rank the semantic groups.

Lastly, the participants were asked to answer a question, that is, when will you use AIM instead of MSN image search. Two scenarios are given that they will prefer AIM. The first scenario is that they are not very familiar with the target images they want to find. Since the first-layer groups display the full aspects of the target, with the user's click-through, the recommendation will guide the user to his required aspect. The second scenario is that the users' target images are hard for them to describe precisely. Under this condition, the interactive recommendation will be very helpful.

4. CONCLUSION

To overcome the common limitations of the existing Web image search engines regarding the passive retrieval process, we proposed an effective mechanism to apply relevance feedback to image retrieval from the World Wide Web. Considering the key issues in Web image search engines when utilizing relevance feedback scheme, namely, real-time requirement, user convention, and user interface, our mechanism shows advantages over traditional methods in three aspects. Firstly, our mechanism regards user's click-through data as implicit relevance feedback to get the user into the loop. Secondly, a hierarchical image search results clustering algorithm is proposed to semantically organize the search results. Using the clustering results as features, our relevance feedback scheme could catch and reflect user's search intension precisely. Lastly, to make the best of the interaction required for reaching good results between user and search engine, a new interface was introduced. Comprehensive user studies show the effectiveness of the proposed implicit relevance feedback mechanism.

5. REFERENCE

[1] Google image search, http://images.google.com

[2] Google web search, http://www.google.com

[3] G.Salton, "Automatic text processing," Addison-Wesley, 1989.

[4] H. J. Zeng, Q. C. He, et al., "How learning to cluster web search results?" *Proc. of the 27th ACM SIGIR*, pp: 210-217.

[5] M. L. Kherfi, D. Ziou, and A. Bernardi, "Image Retrieval from the World Wide Web: Issues, Techniques, and Systems," *ACM Computing Surveys*, vol.36, no.1, pp: 35-67, March 2004.

[6] MSRA clustering search, http://wsm.directtaps.net/

[7] Picsearch image search, http://www.picsearch.com

[8] X. S. Zhou and T. S. Huang, "Relevance feedback for image retrieval: a comprehensive review," *ACM Multimedia Systems Journal, special issue on CBIR*, vol. 8, no. 6, pp: 536-544, April 2003.

[9] Yahoo image search, http://images.search.yahoo.com/