SEARCH RESULT CLUSTERING BASED RELEVANCE FEEDBACK FOR WEB IMAGE RETRIVAL

En Cheng[1], Feng Jing[2], Chao Zhang[3], Lei Zhang[2]

[1]Case Western Reserve University, Cleveland, OH, 44106, U.S.A
[2]Microsoft Research Asia, Beijing, 100080, P.R.China
[3]Huazhong University of Science & Technology, Wuhan, 430074, P.R. China

ABSTRACT

Although relevance feedback (RF) has been extensively studied in the information retrieval community, no commercial Web image search engines support RF because of usability, scalability, and efficiency issues. In this paper, we proposed a search result clustering (SRC) -based RF mechanism for Web image retrieval. The proposed SRC-based RF mechanism employs an effective Search Result Clustering (SRC) algorithm to obtain salient phrases, based on which we could construct an accurate and low-dimensional textual space for the resulting Web images. Given the textual space, we could integrate RF into Web image retrieval in a practical way. The proposed mechanism shows advantage over traditional relevance feedback methods in the following two aspects. On the one hand, our relevance feedback scheme could catch and reflect user's search intension precisely, for the noisy terms would be exempted from the term list with the aid of clustering, thus, the usability of RF in textual space for Web image retrieval is guaranteed. On the other hand, with the exemption of noisy term, the computation with regards to the low-dimensioned textual space is feasible; therefore, the issues of scalability and efficiency for Web image retrieval are addressed. Experimental results on a database consisting of nearly three million Web images show that the proposed mechanism is wieldy, scalable and effective.

Index Terms — Relevance Feedback, Search Result Clustering, Web Image Retrieval

1. INTRODUCTION

With the explosive growth of both World Wide Web and the number of digital images, there is more and more urgent need for effective Web image retrieval systems. Most of the popular commercial search engines, such as Google [3], Yahoo! [9] and AltaVista [1] support Web image retrieval by keywords. There are also commercial search engines dedicated to Web image retrieval, e.g. Picsearch [7]. A common limitation of most of the existing Web image retrieval systems is that their search process is passive, i.e. disregarding the informative interactions between users and retrieval systems. To overcome this limitation, the system should let the user involve into the loop so that personalized results could be provided for specific users. Therefore, there is an urgent need of an effective relevance feedback mechanism applied to image retrieval from the World Wide Web.

Relevance feedback (RF), originally developed for information retrieval is an online learning technique used to improve the effectiveness of the information retrieval system [8]. The main idea of RF is to let the user guide the system. During retrieval process, the user interacts with the system and rates the relevance of the retrieved documents, according to his/her subjective judgment. With this additional information, the system dynamically learns the user's intention, and gradually presents better results. Note that the textual features, on which most of the commercial search engines depend, are extracted from the file name, ALT text, URL and surrounding text of the images. The usefulness of the textual features is demonstrated by the popularity of the current available Web image search engine.

To the best of our knowledge, no commercial Web image search engine supports RF in textual space because of usability, scalability, and efficiency issues. Considering that straightly using the textual information to construct the textual space should lead to a time-consuming computation and the noisy terms will make a negative effect on the performance. Since the user is interacting with the search engine in real time, the RF mechanism should be sufficiently fast, and if possible avoid heavy computations over millions of retrieved images. If only employing Rocchio's algorithm [5], one of the most effective RF algorithms in information retrieval, to Web image retrieval in textual space cannot guarantee a sound performance.

In this paper, we proposed an efficient and effective mechanism to address the above issues. The proposed mechanism employs an effective search result clustering

(SRC) algorithm to obtain the textual features of the resulting Web images. Based on these features, we could construct an accurate and low-dimensional textual space for the resulting Web images. There are two benefits of this procedure. On the one hand, our relevance feedback scheme could catch and reflect user's search intension precisely, for the noisy terms would be exempted from the term list with the aid of clustering. Therefore, the usability of RF in textual space for Web image retrieval is guaranteed. On the other hand, with the exemption of the noisy terms, the real-time requirement of RF based on the low-dimensional textual space is feasible; therefore, the issues of scalability and efficiency for Web image retrieval are addressed.

The remainder of this paper is organized as follows. In section 2, we describe the SRC-based RF mechanism. Experimental results are presented and analyzed in section 3. Finally, we conclude and discuss future work in section 4.

2. SRC-BASEED RF MECHANISM

2.1. Image Representation

To construct our evaluation dataset, near three million images were crawled from several photo forum sites, e.g. photosig 0. These images have rich metadata such as title, category, photographer's comment and other people's critiques. For example, a photo of photosig [1] has the following metadata.

- Title: *early morning*
- Category: *landscape, nature, rural*
- Comment: *I found this special light one early morning in Pyrenees along the Vicdessos river near our house...*
- One of the critiques: *wow...I like this picture very much..I guess the light has to do with everything..the light is great on the snow and on the sky (strange looking sky by the way)...greatly composed...nice crafted border...a beauty*

All the aforementioned metadata is used as the textual source for the following textual space construction, which is based on the Search Result Clustering (SRC) algorithm. The detailed description of the SRC algorithm is illustrated in Section 2.2.

2.2. SRC-based Textual Space Construction

To construct the textual space, we use the SRC algorithm proposed in [4]. The author reformalizes 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 five properties for each phrase. The five properties consist of Phrase Frequency/Inverted Document Frequency (*TFIDF*), Phrase Length (*LEN*), Intra-Cluster Similarity (*ICS*), Cluster Entropy (*CE*), and Phrase Independence (*IND*). The five properties are supposed to be relative to the salience score of phrases. In our case, the comment and critiques are regarded as snippets. In the following, the current phrase (an *n*-gram) is denoted as w, and the set of documents that contains w as $D(w)$. Then, the five properties can be given by

$$TFIDF = f(w) \cdot \log \frac{N}{|D(w)|} \tag{1.1}$$

$$LEN = n \tag{1.2}$$

$$ICS = \frac{1}{|D(w)|} \sum_{d_i \in D(w)} \cos(d_i, c) \tag{1.3}$$

$$c = \frac{1}{|D(w)|} \sum_{d_i \in D(w)} d_i \tag{1.4}$$

$$CE = -\sum_{t} \frac{|D(w) \cap D(t)|}{|D(w)|} \log \frac{|D(w) \cap D(t)|}{|D(w)|} \tag{1.5}$$

$$IND = \frac{IND_l + IND_r}{2} \tag{1.6}$$

$$IND_l = -\sum_{t=l(W)} \frac{f(t)}{TF} \log \frac{f(t)}{TF} \tag{1.7}$$

where

f represents frequency calculation.

Given the above five properties, we could use a single formula to combine them and calculate a single salience score for each phrase. In our case, each term x can be a vector x=(*TFIDF, LEN, ICS, CE, IND*). A regression model learned from previous training data is then applied to combine the five properties into a single salience score y. According to the average performance of linear regression, logistic regression, and support vector regression in [4], the performance of linear regression is the best one. Therefore, in our experiments, we choose the linear regression model. The linear regression model postulates that:

$$y = b_0 + \sum_{j=1}^{p} b_j x_j + e \tag{1.8}$$

where

e is a random variable with mean zero,

b_j is a coefficient determined by the condition that the sum of the square residuals is as small as possible.

The phrases are ranked according to the salience score y, and the top-ranked phrases are taken as salient phrases. The resulting salient phrases constitute the term list, based on which we construct the textual space.

To represent the textual feature, vector space model 0 with TF-IDF weighting scheme is adopted. More specifically, the textual feature of an image I is an L dimensional vector and can be given by

$$\overrightarrow{F} = (w_1, ..., w_L) \tag{1.9}$$

[1] http://www.photosig.com/go/photos/view?id=733881

$$w_i = tf_i \cdot \ln(N / n_i) \quad (1.10)$$

where

$\vec{F}$ is the textual feature of an image I,
w_i is the weight of the i^{th} term in I's metadata,
tf_i is the frequency of the i^{th} term in I's metadata,
L is the number of all distinct terms from clustering results,
N is the total number of images,
n_i is the number of images whose metadata contains the i^{th} term.

2.3. RF in Textual Space

To perform RF in textual space, Rocchio's algorithm [5] is used. The algorithm was developed in the mid-1960's and has, over the years, been proven to be one of the most effective RF algorithms in information retrieval. The key idea of Rochhio's algorithm is to construct a so-called optimal query so that the difference between the average score of a relevant document and the average score of a non-relevant document is maximized. Cosine similarity is used to calculate the similarity between an image and the optimal query. The optimal query can be given by

$$\overrightarrow{F_{opt}} = \overrightarrow{F_{ini}} + \frac{\alpha}{N_{Rel}} \sum_{I \in Rel} \overrightarrow{F_I} - \frac{\beta}{N_{Non-Rel}} \sum_{J \in Non-Rel} \overrightarrow{F_J} \quad (1.11)$$

where

$\overrightarrow{F_{ini}}$ is the vector of the initial query,
$\overrightarrow{F_I}$ is the vector of a relevant image,
$\overrightarrow{F_J}$ is the vector of a non-relevant image,
Rel is the relevant image set,
Non-Rel is the non-relevant image set,
N_{Rel} is the number of relevant images,
$N_{Non-Rel}$ is the number of non-relevant images,
α is the parameter controlling the relative contribution of relevant images and the initial query.
β is the parameter controlling the relative contribution of non-relevant images and the initial query.

In our case, only relevant images are available for our proposed mechanism, so we set α to be 1 and β to be 0 currently.

3. EXPERMIENTAL RESULTS

To construct the evaluation dataset, near three million images were crawled from several photo forum sites, e.g. photosig 0. To automatically evaluate our proposed SRC-based RF mechanism, an image subset was selected and manually labeled as follows. First, ten representative queries were chosen. Then, for each query, the key terms related to the top 20 images were identified. Finally, all resulting images of each query were manually annotated with the corresponding key terms. The key terms and number of result images for each query are shown in Table 1. There are totally 160 key terms.

Table 1. Queries and corresponding key terms. The number within parentheses is the number of result images.

Query	Key terms
Eagle (3809)	creek, eyes, people, place, plant, sunrise, sunset, valley, bald, beautiful, fishing, flying, gray, landing, sea, shout, young
Eiffel tower (1517)	base, diamond, lights, moon, Paris, river, sky, sunset, tier, dark, first, flying, glowing, gothic, into, middle, night, rainy, red, sparking, top, typical, underside, up
Forest house (572)	animal, boat, bridge, flower, nature, lake, people, plant, street, style, autumn, snow
Greek (2646)	beach, building, church, coffee, farmer, goddess, island, light, man, nature, sculpture, ship, street, style, woman, sunset, white
Jaguar (337)	logo, people, racing, type, abstract, classic, e type, old, wild, x-type, animal, car, cat
Merry Christmas (5266)	candle, card, children, gift, light, music, night, ornament, Santa Claus, snow, tree, red, sparking, white
Pear (813)	animal, apple, blossom, leaf, shadow, tree, inside, pair, red
Rainbow (5376)	animal, beach, bird, bridge, falls, horse, light, lorikeet, people, plant, reflection, storm, sunset, valley, double, full, under
Tiger (3826)	butterfly, cat, cub, eye, flower, lily, people, Amoer, blue, common, dark, drinking, plain, Siberian, sitting, sleeping, small, Sumatran, swimming, white, yawning, young
Tulip (3743)	bud, field, people, proportion, blossom, colorful, dry, inside, pink, purple, red, white, yellow

To simulate the interactions between the user and a Web image retrieval system, for each query Q_i, each related key term T_i was selected in turn to represent user's search intention. Images annotated with the term T_i were considered to be relevant to T_i. For each T_i, 5 iterations of user-and-system interaction were carried out. Given a query Q_i, the system first uses the SRC algorithm to build Q_i's textual space S_i. Then, based on the textual space S_i, the relevance feedback using the optimal query learned in equation (1.11) is implemented. After re-ranking the initial resulting images, the system examined the top 20 images to collect the relevant images. Those relevant images labeled in previous iterations were directly placed in top ranks and excluded from the examining process. Precision is used as the basic evaluation measure. When the top 20 images are examined and there are N_{Rel} relevant images, the precision within top 20 images is defined to be $P(20) = N_{Rel} / 20$.

In our experiment, two RF strategies were evaluated and compared: traditional RF and the proposed SRC-based RF. Both of them use Rochhio's algorithm to construct a so-called optimal query. The difference lies in constructing the textual space for the resulting images. Traditional RF uses all terms present in the metadata to construct the textual space, while the SRC-based RF uses the SRC algorithm to obtain the salient phrases, based on which the textual space is constructed. Figure 1 shows the detailed RF performance of the two strategies for the ten representative queries and the average. The average precision of the traditional RF and the SRC-based RF is 0.5481 and 0.6478 respectively. From the result, it can be seen that the SRC-based RF clearly outperforms the traditional RF strategy. The main reason is that using SRC could effectively detect and remove those unimportant or noisy words so that the resulting feature could reflect user's search intension more precisely.

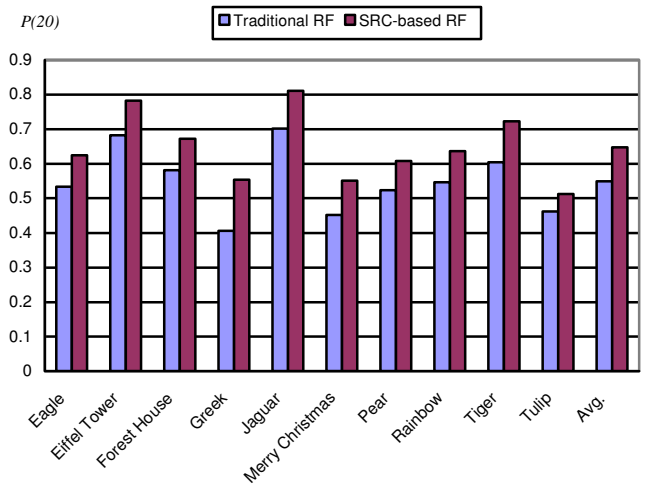

Figure 1. Precision comparison of two RF Strategies

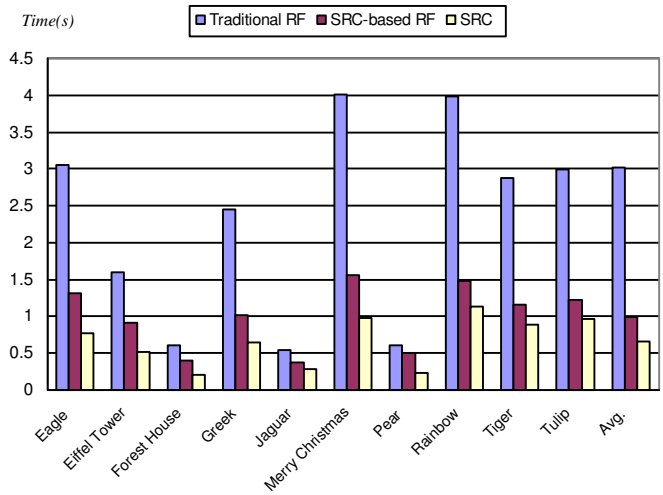

Figure 2. Efficiency comparison of two RF Strategies

Besides the performance comparison, the time cost of the two strategies is another factor worth analyzing. Given a query Q_i and a term T_i, the time cost for completing 5 iterations of user-and-system interaction is recorded. Based on the sum of each term's time cost, we could obtain the average time cost for each query Q_i. For SRC-based RF, we also record the average time cost for accomplishing the SRC procedure. Note that each query need accomplish only one SRC procedure, and the resulting textual space is suitable for all the related terms. Figure 2 shows the time cost of the two strategies for the ten representative queries and the average. The average time cost of the traditional RF, SRC-based RF, and SRC is 3.02s, 0.994s, and 0.664s respectively. From the result, it can be seen that the SRC-based RF mechanism is more efficient than the traditional RF. Therefore, the SRC-based one is more practical for a real Web image retrieval system.

4. CONCLUSION

In this paper, we proposed a search result clustering based relevance feedback mechanism for Web image retrieval. The proposed mechanism employs an effective search result clustering algorithm to obtain salient phrases, based on which we could construct an accurate and low-dimensional textual space for the resulting Web images. As a result, we could integrate RF into Web image retrieval in a practical way. Experimental results on a database consisting of nearly three million Web images show that the proposed mechanism is wieldy, scalable and effective.

Besides explicit relevance feedback, implicit relevance feedback, e.g. click-through data [2] can also be integrated into the proposed mechanism. Moreover, since only textual features are used in the proposed algorithm, other Web media, e.g. music or video could also be retrieved with the proposed RF algorithm as long as they have textual descriptions.

5. REFERENCES

[1]Altavisa image search, http://www.altavista.com/image/

[2]E. Cheng F. Jing, L. Zhang, and H. Jin, "Scalable Relevance Feedback Using Click-through Data for Web Image Retrieval," *Proc. of the 14th annual ACM international conference on Multimedia*, pp.173-176, 2006.

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

[4]H. J. Zeng, Q. C. He, Z. Chen, W. Y. Ma and J. W. Ma, "Learning to cluster Web search results," *Proc. of the 27th annual international ACM SIGIR conference*, pp. 210-217.

[5]J. Rocchio, Relevance Feedback in Information Retrieval. The SMART Retrieval System Experiments in Automatic Document Processing, Prentice Hall, pp.313-323, 1971.

[6]Photosig, http://www.photosig.com

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

[8]R. Baeza-Yates, and B. Ribeiro-Neto, "Modern Information Retrieval". Addison-Wesley, June 1999.

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