Human Behaviour Consistent Relevance Feedback Model for Image Retrieval

Jing Liu
Institute of Automation
Chinese Academy of Sciences
Beijing 100080, China
+86-10-62542971
jliu@nlpr.ia.ac.cn

Zhiwei Li, Mingjing Li
Microsoft Research Asia
49 Zhichun Road
Beijing 100080, China
+86-10-58968888
{mjli, zli}@microsoft.com

Hanqing Lu, Songde Ma
Institute of Automation
Chinese Academy of Sciences
Beijing 100080, China
luhq@nlpr.ia.ac.cn,
mostma@gmail.com

ABSTRACT
Due to the well known semantic gap, content based image retrieval is a difficult problem. To bridge it, relevance feedback as an effective solution has been extensively studied in literatures. However, existing methods follow a single-line searching philosophy, which may lead to a local optimum in search space. To address the problem, we propose a human behavior consistent relevance feedback model for image retrieval in this paper. Simulating human behaviors, the proposed model enable the user to perform relevance feedback in three manners: *Follow up*, *Go back*, and *Restart*. Each manner is a way for the user to provide the system with his or her opinions about search results. The accumulated feedback information can be used to refine the user query and regulate the similarity metric. We adopt the graph ranking algorithm to model the retrieval process. Experiments conducted on standard Corel dataset and Pascal VOC 2006 dataset demonstrate the effectiveness of the proposed mechanism.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Retrieval Models

General Terms
Algorithms, Measurement, Experimentation

Keywords
Image retrieval, relevance feedback, graph ranking.

1. INTRODUCTION
Content based image retrieval is a hot research topic. However, this problem is very difficult due to the well known semantic gap. Basically, the difficulties can be viewed from two aspects. First, although we can extract lots of visual features from an image, we do not know which features are coherent with semantics of the image. Second, a query image can have many semantics from different views, which one is the user's intention? Under this background, relevance feedback (RF) is introduced to involve the user's interaction to address these issues.

MM'07, Sep 23–28, 2007, Augsburg, Bavaria, Germany

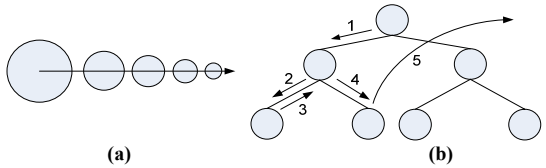


Figure 1. Illustrations of two types of search processes: (a) Traditional single-line one; (b) Human behavior consistent one, in which steps 1, 2, and 4 indicate "Follow up", step 3 denotes "Go back", step 5 denotes "Restart".

Typically, image retrieval with RF adopts an iterative manner. At each round of the iteration, the system gives the user some images, and expects the user to give relevant or irrelevant judgment to each image. Once the system gets feedbacks from the user, it can re-rank images in database to select more relevant images for the user. Indeed, this process can be deemed as an iterative process to search over image space to find relevant images. As illustrated in Fig. 1(a), we are in a big space at the beginning. In the following steps, we struggle to narrow down the space until the user finds what he or she needs. Narrower and narrower is a good description to this process because the user can never return to a bigger concept in search. This is right the philosophy existing methods follow. However, is this the right way for people to search information? Definitely, the answer is no because the user may step into an over-narrow space or even a wrong space in the single-line search process.

For a common user, a search process is actually a complex travel on the underlying hierarchical semantic tree rather than simply along a single-line path. Based on this view, we design a hierarchical retrieval process as illustrated in Fig. 1(b), in which three types of user behaviors in search are considered: *Follow up*, *Go back*, and *Restart*. Each type of behaviors can be deemed as a kind of implicit feedback to provide the system with the user's opinions about search results. Besides, the relevance judgments from the user are required as a kind of explicit feedback.

With the feedback information, the system dynamically learns the user's intention, and gradually presents better results to adapt to the user. Currently, studies of machine learning on the adaptation have been received much attention. State-of-the-art techniques can be classified into inductive and transductive ones. The goal of an inductive method is to create a classifier which separates the relevant and irrelevant images and generalizes well on unseen examples. One typical approach is based on SVM [5]. However, the major problem is the insufficiency of labeled examples, which may weaken the classification performance. In contrast to

inductive methods which only rely on the labeled set, transductive approaches integrate the unlabeled data. A representative work is the graph ranking based method [2], which is most similar to our work. The method evaluates the relevance between two images by a fixed similarity measure over the original content feature space and performs image retrieval in a single-line manner.

In this paper, a human behavior consistent relevance feedback model is proposed, which allows the user to perform three kinds of feedback operations, as well as explicit relevance judgments. Obtaining all these feedbacks, we adopt a transductive learning method, i.e., the graph ranking algorithm, to model this process. At each round of feedback, three main tasks would be performed in the process. First, the user's feedbacks will be accumulated to the label vector in a straightforward manner. Second, the label information and those feedbacks will supervise us to adjudge the query and further to learn a better distance metric. Finally, the label information will be propagated in the whole database based on the similarity graph built with the updated distance metric.

2. IMAGE RETRIEVAL WITH RF

Here, we adopt the graph ranking algorithm to unify the feedbacks and the similarity measure together, while unlabeled data are also incorporated into the transductive learning process. For clarity, we will first introduce the graph ranking algorithm.

2.1 Graph Ranking Algorithm

Given N points $X=\{x_1,...,x_N\}\subset\Re^d$, x_i is a d-dimensional feature vector. The initial label vector is defined as $y=[y_1,...,y_N]^T$. At the beginning, we usually define labels of certain nearest neighbors of the query to 1, and the rest to 0. The item in vector $f\in\Re^N$ is the ranking score of each point related to the query, whose stable result is obtained from the following iterative procedure as in [2].

Step 1: Construct the similarity matrix $W\in\Re^{N^2}$ as:

$$W_{ij}=\exp\left[-dis\left(x_i,x_j\right)/\sigma\right] \tag{1}$$

where $\sigma>0$, $W_{ii}=0$, and $dis(.)$ is certain distance metric.

Step 2: Symmetrically normalize W as:

$$S=D^{-1/2}WD^{-1/2},\quad D_{ii}=\sum_{j=1}^{N}W_{ij} \tag{2}$$

Step 3: Do iteration according to Eq. 3, until convergence.

$$f\left(t+1\right)=\theta\times S\times f\left(t\right)+\left(1-\theta\right)\times y,\quad \theta\in[0,1] \tag{3}$$

where t represents the number of iterations, and $f(0)=y$.

Step 4: Obtain the ranking score of each point according to f^*.

2.2 Basic Flow of Image Retrieval

We take one round of image retrieval as an example to introduce the basic flow of our method, which is illustrated in Fig. 2. Assuming a ranking score vector (f_n^*) has been obtained in the nth round, the operations in the $(n+1)$th round are given as follows:

Feedback Collection: Collect user's feedbacks about the result of f_n^*, which include explicit relevance judgments (r_{n+1}) and one of the three operations (b_{n+1}), i.e. *Follow up*, *Go back*, or *Restart*.

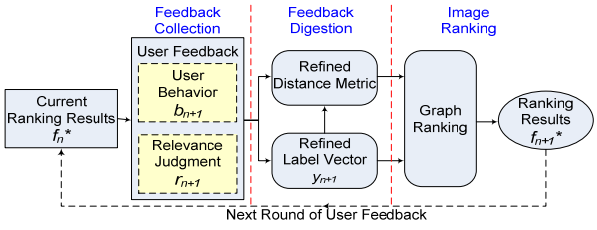


Figure 2. Basic flow of image retrieval.

Feedback Digestion: Two steps will be performed to digest user intentions: 1) Use the feedbacks to update the initial label vector y_{n+1}. 2) Combine the feedbacks with the updated label vector to learn a desired distance metric, i.e. $dis(.)$ as in Eq. 1.

Graph Ranking: Unify the refined label vector and the updated distance metric to perform the graph ranking, and obtain the ranking result (f_{n+1}^*) in this round.

Note that the items in f_{n+1}^* should be scaled into [-1, 1] as Eq. 4, to make the ranking result usable for the next round feedback.

$$f_{n+1}^*(i)\begin{cases}>0 & \text{scaled to (0,1]}\\ =0 & \text{none change}\\ <0 & \text{scaled to [-1,0)}\end{cases} \tag{4}$$

The items in f_{n+1}^* corresponding to the relevant and irrelevant samples are scaled separately, and the neutral ones have no change. This aims to prevent the impact of positive samples from being overwhelmed by negative samples, since they are distributed asymmetrically.

3. HUMAN BEHAVIOR CONSISTENT RF

As discussed above, the initial label vector and the distance metric are two important items to digest user feedbacks as well as to the graph ranking algorithm. In the following, we will focus on the schemes to design these two items in the RF process.

3.1 Quantitative Discrimination of Operations

For analyzing convenience, we first introduce two boundaries to discriminate the three types of implicit feedbacks quantitatively. One is the "tolerable precision", which denotes the lowest tolerable precision of current search results for the user. The other is the "acceptable precision", which denotes the lowest acceptable precision of current search results. Then, we explain the three operations with the two boundaries, in which the tolerable precision, the acceptable precision, and the precision of search results (the nth round) are denoted as λ_1, λ_2, and P_n respectively.

When $\lambda_2\le P_n\le 1$, the user would select "Follow up": The user basically accepts the results. However, he expects better results can be achieved by further feedbacks, and follows up the search.

When $\lambda_1\le P_n<\lambda_2$, the user would select "Go back": The search results are unacceptable and not better than the previous results. Then the user would select to return to the last step without any relevance judgment.

When $0\le P_n<\lambda_1$, the user would select "Restart": The current results are too bad to be tolerated by the user, and he has to restart the search from another entrance without any relevance judgment.

Through above explanation, the three operations implicitly express the user opinions about the search results and indicate different qualities of the results indeed. Accordingly, different schemes to digest the user's feedbacks are necessary.

3.2 Updating Label Vector

According to the user operation, three updating schemes for the label vector are designed and the illustrations are given in Fig. 3.

Follow up ($\lambda_2\le P_n\le 1$): When the user performs the operation, we can reasonably deduce that more than $\lambda_2\times 100\%$ images in

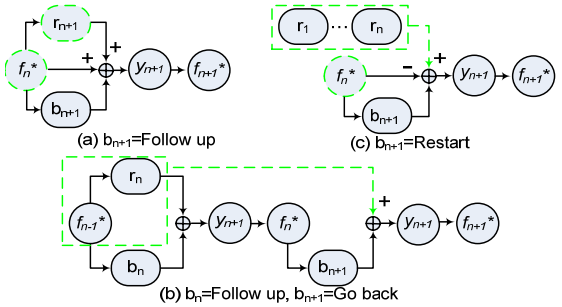


Figure 3. Simplified illustrations for different user behaviors, in which the parts outlined with green dashes are ones contributing to the refined label vector of y_{n+1}.

current results are relevant to the user's query. Then we combine the ranking vector (f_n^*) with the relevance judgments (r_{n+1}) in this round to update the initial label vector y_{n+1} as:

$$y_{n+1} = \lambda_2 f_n^* + r_{n+1} \tag{5}$$

For the elements in r_{n+1}, those of corresponding to identified relevant images are set to 1, those of corresponding to irrelevant images are set to -1, and the rest are set to 0.

Go back ($\lambda_1 \le P_n < \lambda_2$): This operation indicates that a lower quality has been achieved for current search results compared with previous results. This is possibly because too biased relevance judgments or certain unsuitable distance metric are provided in the prior retrieval process. As a conservative solution, we weaken the relevance judgments on the last round (r_n). Besides, since the search result in the (n-1)th round (f_{n-1}^*) has been accepted by user, we can confidently regard the result to be "good". Then the information from r_n and f_{n-1}^* is used to update the label vector as:

$$y_{n+1} = \lambda_2 f_{n-1}^* + \eta r_n \tag{6}$$

where η is a regulating parameter to control the weakened effect. In our implementation, η=0.5.

Restart ($0 \le P_n < \lambda_1$): When the operation has been performed, most of search results, especially those on the first page (30 images in our implementation) are likely to be irrelevant to the given query. As an extreme solution, all the images on the first page except for those having been identified as positive samples are considered as negative samples, while the identified positive ones are weakened manually. The updated label vector is given as:

$$y_{n+1}(i) = \begin{cases} -1 & \text{if image } i \text{ is irrelevant or is ranked on the first page} \\ 0.5 & \text{if image } i \text{ is relevant} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

3.3 Distance Metric Learning

Obtaining the refined label vector, we would seek some fresh knowledge about a "good" sample distribution to enable the effective graph ranking. It requires us to find a reasonable query point and learn an effective distance metric to represent the data distribution. Here, we consider a parametric distance metric as:

$$dis(x,y) = dis_A(x,y) = \|x-y\|_A = \sqrt{(x-y)^T A (x-y)} \tag{8}$$

where $x, y \in \Re^d$, and $A \in \Re^{d^2}$ is used to project each point from its original space to a more discriminative space, and d is the feature dimension. We require A>0 to ensure $dis_A(.)$ to be a metric.

To get a desired metric for image retrieval, our proposal is to use the supervision from the label vector (y_{n+1}) to direct the metric learning. The criterion is that the point in dataset with high rank score should have small distance with the query vector, while negative points should be far away from the query point. The optimization problem is defined as:

$$\min_A \sum_{i=1}^N y_{n+1}(i) \|x_i - q_{n+1}\|_A^2, \quad s.t. \quad \det(A) = 1, \quad A>0 \tag{9}$$

where the constraints are added to prevent A from a trivial value, and q_{n+1} is an updated query vector defined as:

$$q_{n+1} = q + \alpha \cdot \frac{1}{C_1} \cdot \sum_{i=1}^{N_{q^+}} y_{n+1}(q_i^+) \cdot x_{q_i^+} + \beta \cdot \frac{1}{C_2} \cdot \sum_{i=1}^{N_{q^-}} y_{n+1}(q_i^-) \cdot x_{q_i^-} \tag{10}$$

$$C_1 = \sum_{i=1}^{N_{q^+}} y_{n+1}(q_i^+), \qquad C_2 = \sum_{i=1}^{N_{q^-}} y_{n+1}(q_i^-),$$

where q is the initial query vector, $y_{n+1}(q_i^+)$ denotes the (q_i^+)th item of the label vector y_{n+1}, N_{q^+} and N_{q^-} are the numbers of positive and negative samples, and $\alpha, \beta > 0$ control the effects of relevant and irrelevant images (In our experiments, $\alpha=\beta=1$).

In our implementation, only the case that A is a diagonal matrix is considered. The solution with the Lagrange multiplier algorithm is given as Eq. 11, whose detailed proof can be referred to [3].

$$A_{jj} \propto 1/\sigma_j^2, \quad \sigma_j = \sum_{i=1}^N y_{n+1}(i)(x_{ij} - q_{n+1})^2 \tag{11}$$

where i=1,2,...,N, and j=1,2,..., d. In case of σ_j=0, the problem can be solved by the method proposed in [3].

Obtaining the optimized matrix A, the new distance metric (Eq.8) is used to update the similarities in Eq. 1. With the updated similarity matrix and the refined label vector, we can re-rank images in database by the graph ranking algorithm, and prepare search results for the next round feedback. Those top ranked images will be selected as the search results to the user.

4. EXPERIMENTS

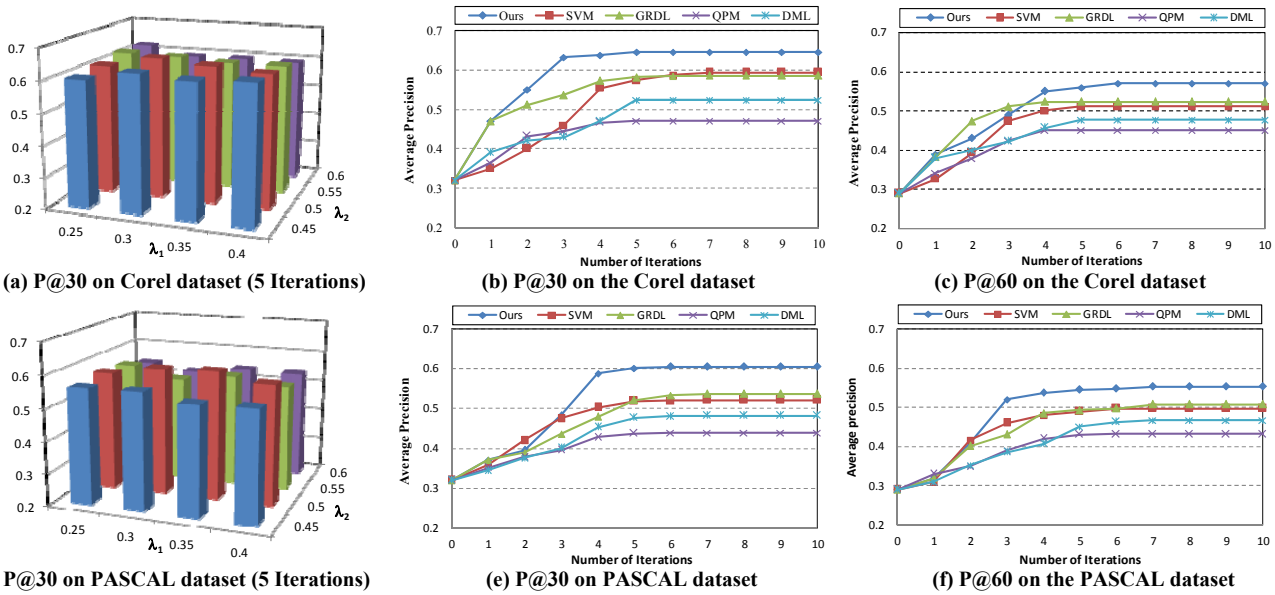
4.1 Experimental Design

In our experiments, two benchmark datasets are used. The first one is the Corel dataset, which includes 5000 color images from 50 categories and each category has 100 images. The second dataset is used in the PASCAL VOC 2006 challenge [1], in which there are 5304 color images. These images distribute in 10 categories. For both datasets, images from the same category are deemed to be relevant to each other. We extract a 224-dimensional visual feature for every image: 36-dimensional color histogram in HSV color space, 144-dimensional correlogram feature, 24-dimensional pyramid wavelet texture feature, and 20-dimensional Tamura feature.

Similar to other previous work on RF, we design an experimental procedure to perform experiments in an automatic manner and evaluate the performance of image retrieval with RF. We randomly sample 300 queries from each dataset. To simulate the user's behaviors at each round of the feedback, we use the two boundaries, i.e. tolerable precision (λ_1) and acceptable precision (λ_2), to decide the user's operations. If the precision is higher than λ_2, the user will take "Follow up"; if the precision is between λ_1 and λ_2, the user will take "Go back"; otherwise "Restart". The initial retrieval results are obtained by ranking images according to the Euclidean distance without any relevance feedback. Average precision of top m search results over the 300 queries, denoted as P@m, is used to evaluate the performance.

4.2 Discussion on Two Boundaries

The two boundaries of λ_1 and λ_2, can be used to quantify the satisfaction of the user to the search result and infer his or her intention in search. Fig. 4(a) and Fig. 4(d) show the experimental results on both dataset with five iterations of RF. When we range



(a) P@30 on Corel dataset (5 Iterations)
(b) P@30 on the Corel dataset
(c) P@60 on the Corel dataset
(d) P@30 on PASCAL dataset (5 Iterations)
(e) P@30 on PASCAL dataset
(f) P@60 on the PASCAL dataset

Figure 4. Results on the two datasets.

λ_1 from 0.25 to 0.40, and range λ_2 from 0.40 to 0.60, the average precisions (*P@30*) change slowly. That is, the retrieval performance is not sensitive to both boundary precisions, although they seem to be different from person to person and may affect the performance. This makes the design of such a retrieval system based on the proposed model feasible and robust. Here, we set λ_1 to 0.30 and λ_2 to 0.50 as defaults.

4.3 Comparisons with Other Algorithms

We compare our algorithm with other four algorithms: SVM [5], GRDL (Graph Ranking [2] adding to our proposed Distance Learning), QPM (Query Point Movement) [4], and DML (Distance Metric Learning) [3]. Since the four related algorithms are performed as the single-line search, we only take "Follow up" operation at each round of feedback for them. For SVM, we adopt RBF kernel; for QPM, the two coefficients to control the roles of positive and negative samples in query updating are set to 1; for GRDL and our method, the parameters in graph ranking algorithm are set as $\theta=0.99$ and $\sigma=0.05$, which are consistent with [2].

Fig. 4 (b, c) and Fig. 4(e, f) show the average precisions on the two datasets respectively. Our algorithm significantly outperforms the other four algorithms. Among these baseline algorithms, GRDL is most similar to our method except that it does not allow the "Go back" and "Restart" operations. The lower performance of GRDL than ours indicates the necessity of the two additional operations. It can also be observed that our method by allowing the additional feedbacks enables a higher ascending speed to achieve the better performance than other methods. In addition, we count the numbers of the two unique operations, i.e. "Go back" and "Restart", performed in the experiment. In five iterations of image retrieval with RF, the "Go back" operation is performed 0.63 times averaged over 600 queries, which include 300 queries for Corel dataset and 300 for PASCAL dataset. The case for the "Restart" operation is 0.29 times. The occurrences of both additional operations lead the search to a correct direction almost in less than 5 iterations.

In all, we can safely conclude that the really encouraging issues in our proposed retrieval model are the three types of relevance feedbacks consistent with human behaviors, while the two boundary precisions only need to be around a suitable range.

5. CONCLUTIONS

In this paper, we propose a human behavior consistent RF model, in which human behaviors are regarded as three kinds of implicit feedbacks. At each round of feedback, we firstly embed the feedback information into the representation of the label vector, and learn a new distance metric supervised by the feedbacks. Then, we unify both items to re-rank images in database by using the graph ranking algorithm. The experimental results indicate: i) the three kinds of operations are very necessary to capture the user's intentions in search, especially the new operations (*Go back* and *Restart*) are preferable to lead the search to a correct direction; ii) the proposed learning method and the ways to digest the user's feedback are more effective than some related work.

6. ACKNOWLEDGEMENTS

The research was supported by National 863 Project (2006AA01Z315), National Natural Science Foundation of China (60121302), and Beijing Natural Science Foundation (4072025). This work was performed at Microsoft Research Asia.

7. REFERENCES

[1] M. Everingham, A. Zisserman, C. Williams, L.V. Gool. *The PASCAL visual object classes challenge 2006*. In 2th PASCAL Challenge Workshop.

[2] J. He, M. Li, H.J. Zhang, H. Tong, and C. Zhang, *Manifold-Ranking Based Image Retrieval*, Proc. ACM International Multimedia Conference, 2004.

[3] Y. Ishikawa, R. Subramanya, C. Faloutsos. *MindReader: Querying Database through Multiple Examples.* New York, USA: Proc. the 24th VLDB Conference, 1998.

[4] J. Rocchio. *Relevance feedback information retrieval.* Gerard Salton (ed.): The Smart Retrieval System-Experiments in Automatic Document Processing, pp. 313–323. Prentice-Hall, Englewood Cliffs, N.J., 1971.

[5] L. Zhang, F. Lin, and B. Zhang. *Support vector machine learning for image retrieval.* ICIP, pp. 721-724, 2001.