AN ADAPTIVE-WEIGHT HYBRID RELEVANCE FEEDBACK APPROACH FOR CONTENT BASED IMAGE RETRIEVAL

*Yi Zhang, Wenbo Li, Zhipeng Mo, Tianhao Zhao, Jiawan Zhang**

School of Computer Software, Tianjin University

ABSTRACT

Content-based image retrieval (CBIR) has been receiving intensive research attention for many applications. In order to provide the users with more precise retrieval results, relevance feedback (RF) methods have been incorporated into CBIR which take the user's feedbacks into account. In general, explicit RF methods demand too much user effort while implicit RF methods suffer from lower retrieval accuracy. As such, we propose a hybrid RF method, *adaptive-weight hybrid relevance feedback* (AHRF) for content-based image retrieval. AHRF integrates explicit user grading and implicit user browsing histories to build a user preference model. The model is refined iteratively and used to train a preference classifier for the users. Moreover, an adaptive-weight mechanism is proposed to achieve a personalized preference model. Our proposed method is tested on a subset of the Corel Database and the experimental results reveal that AHRF can achieve good retrieval precision with less user effort.

Index Terms— CBIR, Relevance feedback, hybrid, adaptive weight

1. INTRODUCTION

[1]The need of image retrieval grows quickly with the number of images on the Internet increasing explosively. The traditional annotation based image retrieval depends heavily on the manual descriptions [1,2] for the images such as file-names, categories, annotated keywords. Unfortunately, high-priced manual annotation and inappropriate automated annotation are always frustrating users in image retrieval applications. To deal with such problems, researchers turned to content-based image retrieval (CBIR), which has been currently an active research topic.

One difficulty causing retrieval accuracy reduction in CBIR is the semantic gap between image visual features and user understanding. To obtain more precise retrieval results, relevance feedback (RF) techniques [3] are incorporated into CBIR by taking user's feedbacks into account in the next retrieval process. Usually, RF methods require the users evaluate many images in multiple iterations to find out users' real intents. Therefore, there exists the contradiction between the retrieval accuracy and the required user's effort in the RF methods. To resolve this contradiction, much work has been done. To our best knowledge, there are mainly two types of strategies so far: reducing the number of iterations and reducing the users' burden in each iteration.

For the former strategy, Su et al. [4] proposed Navigation-Pattern-based Relevance Feedback (NPRF) to achieve high retrieval quality of CBIR with RF by using discovered navigation patterns. Their work aims at reducing the redundant browsing and reaching exploration convergence quickly by mining the user query logs. They report that their method can achieve accuracy of over 90% in 6 iterations. As most RF methods, NPRF also suffers from the drawbacks of heavy users' burden. And the retrieval accuracy relies on users' active involvement, so that the faulty operations may lead to performance reduction.

As for the latter strategy, the implicit RF methods are proposed. Different with the explicit RF, the implicit RF technique [5] gathers useful data indirectly by monitoring behaviors of the users during the searching and browsing process instead of requiring much effort from the users. This kind of technique was first applied in retrieving documents and was brought into CBIR several years ago. A number of studies that employ implicit RF have been made to reduce the users' burden. Auer et al. [6] proposed a system which infers users' intent from eye movements by using a machine learning method. Then the system learns a similarity metric of common image features depending on the current interests of the user. Similarly, the system of [7] improves the performance of image retrieval by re-ranking the retrieved images according to color and texture features extracted from the regions where the users pay more attention. The users' interests are found by gazing information collected from an unobtrusive eye tracker. Although the implicit RF methods are quite profitable in liberating the user from heavy burdens, they are generally thought to be less accurate than explicit RF methods [8].

By combining explicit RF and implicit RF, Zhang et al. [9] introduced a user-driven model to improve retrieval accuracy of the implicit RF method. They studied users' browsing habits and built a preference model to re-rank the retrieved images. However, they use the uniform parameters in the model, and thus cannot deal with the problem of user personality.

* Corresponding author: jwzhang@tju.edu.cn

As an improvement, we propose *adaptive-weight hybrid relevance feedback* (AHRF) method in this paper. AHRF integrates explicit grading and implicit browsing habits, building the personalized preference model and achieving good retrieval precision with less user effort in an iterative way. In each iteration, the user is required to grade a portion of the images; meanwhile four types of data are collected implicitly: browsing time, download, scrolling and zoom-in. For the graded images, the user's preference values are recorded as the grades; for the images without explicit grades, the user's preference values are automatically inferred according to the user's behaviors and the explicit grades for other images. Then the preference values will be used to train an image classifier to create the image set to be browsed for the next iteration. The above process will be iterated several times to get the stable preference values for the retrieval results. Considering that different users may perform quite variously during the browsing process, we also propose an adaptive-weight mechanism to solve the personality problem. Our method is tested on Corel image set and the experimental results show that our method could achieve good retrieval precision.

The rest of the present paper is organized as follows. Section 2 describes both the framework and the technical details of our method. In Section 3, we provide and evaluate the experimental results. Finally section 4 concludes the paper.

2. METHOD

2.1. Overview

The steps of AHRF method are summarized as follows:

1) Given an unordered initial retrieval image set of the user's query, the user is required to browse and grade images. The system implicitly records the user's manipulations.
2) The preference value of each image is calculated. For graded images, their preference values are recorded as the user's grades. Otherwise, the preference values are estimated by the adaptive-weight preference model built on users' grading and manipulation histories.
3) According to the preference values, the browsed images are divided into a positive group and a negative group. Thereafter, three types of image features, which are color, edge and texture, are extracted for these images and used as the support vectors in training an SVM classifier. Then the retrieved images are reordered by giving priority to those positive classified images.
4) The reordered image set is provided to the users for grading once again. A stable preference model for each user is achieved after a few iterations.
5) The preference model and the classifier can be reused for the same user when receiving the similar queries.

2.2 Adaptive-weight preference model

In the adaptive-weight preference model, we assume that each image has a preference value that reflects the degree of user's satisfaction. The retrieval results will be ranked in descending order by preference values. Thus the target of the model is to assign the proper preference values to the images.

When building the model, we integrate the explicit and the implicit relevance feedbacks. For the explicit part, the users are required to grade a percentage of images when browsing the retrieval results. As for the implicit part, four kinds of user's operations are considered: download, browsing time, scrolling and zoom-in. By analyzing the users' browsing behaviors, we find out a trend that the downloaded images are likely to be given high grades. Also, the images with long browsing time, multiple scrolling and zoom-in operations mean that they get more attention from the users.

According to the users' behavior, the initial images given to the users can be classified into browsed images and un-browsed images. We only collect the users' behaviors of browsed images because un-browsed images are totally ignored by the users. Having the records of relevance feedbacks, we calculate the preference value for each browsed image via Eq. (1).

$$V(I)=\begin{cases} H^{d}[\sum_{i=1}^{3} w_i * S_i(I)]^{|d-1|} & , G(I)=0 \\ G(I) & , G(I)>0 \end{cases} \quad (1)$$

In the above formula, $V(I)$ is the preference value of browsed image I and $G(I)$ is the grade given for I in the explicit feedbacks. d is a binary value which represents whether the image is downloaded or not. Other three implicit factors, browsing time, scrolling count and zoom-in count are denoted by $S_i(I)$ (i = 1, 2, 3) with corresponding weights w_i (i = 1, 2, 3) respectively. H is the highest grade the user can give to an image. In our method, we restricted the grade to integer from 0 to 5, thus having $H=5$ and $G(I)=0$ for those images which are not graded or browsed.

Our method requires that the user browses a percentage such as 50% of the images, and grade lower percentage of the images. For the graded images, their preference values can be computed in two ways, one is the grades given by the users and the other is the weighting implicit factors. For those images with only implicit feedbacks, their preference values can be calculated using the same weights w_i (i = 1, 2, 3) as graded images for one user.

For different users, the weights vary with each other. In order to estimate w_i (i = 1, 2, 3) dynamically, a group of equations are built for each user as follows:

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ \dots & \dots & \dots \\ S_{m1} & S_{m2} & S_{m3} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} V_1 \\ V_2 \\ \dots \\ V_m \end{bmatrix}, \quad (2)$$

where m is the number of graded images. Furthermore, we get the simplified form $Sw = V$. The least square method (LSM) is adopted to solve optimal w_i (i = 1, 2, 3) by minimizing the square errors:

$$\min_{w} \| Sw - V \|_2, \quad S \in C^{m\times 3}, \quad V \in C^3 \quad (3)$$

Though responsible participations of the users are encouraged, we can't guarantee every feedback is valid. To reduce the impact of faulty operations, we adopt average value of the quasi optimum values as final weights instead of directly using optimal values. In brief, the procedure can be decomposed into several steps as follows:

1) The simultaneous equations are divided into n subgroups satisfying the conditions that each subgroup includes the equations from all grade levels. Then n subgroups of equations are built: $S_j w = V_j (j=1,2\dots n)$. We use n=5 in this paper.
2) The optimal solution $w_{i,j}$ (i = 1, 2, 3, j=1,2…n) for each subgroup of equations are calculated using LSM.
3) Finally, we take the average values $\overline{w}_i (i = 1,2,3)$ as the final solution via Eq. (4).

$$\overline{w}_i = \frac{\sum_{j=1}^{m} w_{i,j}}{n}, \qquad i = 1,2,3 \quad (4)$$

Having estimated $\overline{w}_i (i = 1,2,3)$, we use them to calculate the preference values of browsed images which are not explicitly graded by the user. Then the browsed images are classified into the positive group and the negative group at a threshold of preference value. The threshold is 3 in our paper. Then a preference classifier is trained to classify the images which have not been browsed by the user.

2.3 Feature extraction and training

We extract three image features for training, which are color histogram, edge and texture.

The color histogram is counted in HSV color space with 32 bins, 16 bins, and 16 bins for H channel, S channel and V channel respectively. Therefore, a 64-dimension color feature vector is obtained for each image.

The edge feature is extracted in gray color space. Firstly we convert the color images to grayscale images and use canny operators to get the edge of images. Then the edge images are divided into 256 regions, calculating the ratio of edge pixels in each region. This 256-dimension vector is used as the edge feature.

The texture feature extraction is also performed on grayscale images. We divide the gray images into partially overlapping regions by 48*48 pixels. Then 2D Gabor filter is applied to each region. Basically, the importance of the texture is related to the area size. That is, the texture coving more regions is considered more important. Thus in order to extract primary textures, we adopt K-means to cluster similar textures together and select some top representative textures as the texture features for the image.

For positive images and negative images, all three types of the features are extracted as the support vectors in SVM training. Then the un-browsed images are classified using the trained preference classifier. With the browsed images, all classified images are reordered and showed to the user for the next feedback iteration.

2.4. Iterative relevance feedback

Our proposed AHRF is an iterative method. In each iteration, the system presents the reordered retrieval results and receives the feedbacks from the user once again. On the basis of refined feedbacks, the adaptive-weight preference model is rebuilt and new preference classifier is created. After several iterations, the stable preference model and the preference classifier are obtained for each user. As there is no special browsing requirements such as non-overlapping rule in later iterations, the user tend to repetitively grade very satisfying images and some fresh satisfying images. The classification accuracy will stop improving when no new grading images are added, thus our method converges very fast.

3. EXPERIMENTS AND RESULTS

To evaluate the effectiveness of our method, we conduct three experiments. First, the accuracy of adaptive-weight estimation is tested. Second, the overall performance and convergence of AHRF is evaluated. Finally, we compare the performance of AHRF under different degrees of user engagement.

The experimental data comes from the Corel image database. We prepared 6 data sets composed of 15 different categories. The expected retrieval targets of 6 data sets are butterfly, mountain, woolwork, flower, sail boat and water wave. The original categories include butterfly, bird, deer, cat, mountain, flower, sail boat, sea, water wave, stones, grass, tree bark, grain, marble and woolwork. More details of the data can be found in Table.1.

3.1 Evaluation of adaptive-weight preference estimation

In measuring the accuracy of the proposed adaptive-weight preference estimation, we ask the user to grade all experimental images, and then the estimated preference values are compared with the ground-truth. The accuracy is defined by ratio between correctly estimated images and all the images. As mentioned, the images are classified into positive group and negative group by preference value at

threshold 3. Furthermore, the estimation is regarded as correct if it can be classified to the same group with value error within 1.

In order to evaluate the estimation accuracy of average sub-optimum in the noisy condition, we randomly add about 15% fault grading records into users' operations. The accuracy comparison of absolute optimal estimation and average sub-optimum estimation is showed in Fig. 1. The comparison reveals that the average sub-optimum estimation is more stable than absolute optimal estimation and can achieve high accuracy for most data sets.

3.2 Evaluation of performance and convergence

To evaluate the effectiveness of proposed approach, the commonly used performance metrics, precision and recall, are used. Precision is defined as the ratio of correctly retrieved images to all retrieved images, and recall is defined as the ratio of correctly retrieved images to all relevant images should be retrieved.

Fig.2 illuminates that the precision rises with the iteration proceeding. The results show that AHRF reaches the stable precision within only two iterations. The best precision and recall for each data set is shown in Table 1. As reported, implicit RF method Pinview [6] achieves the average precision 0.224 for a subset of the Corel image database. And an adaptive explicit RF method FARF [10] achieves the average precision over 50%, average recall about 10% for a subset of the Corel database containing 1400 images from 14 categories. By comparing with them, we can generally conclude that our method can effectively achieve good retrieval performance. In the future, the extended experiments on larger dataset are needed to verity the advantage of our method.

3.3 Evaluation of user engagement

In addition to the effectiveness, another issue we are interested in is the user efforts needed in our method. Usually, our method expects the users browse at least 50% images and grade at least 25% images in each iteration. However, that is not a mandatory requirement. We test our method assuming that the users reduce their participations in grading process. Fig. 3 exhibits that our method is not very sensitive to negative user engagement, though more grading operations can improve the retrieval performance.

Table1 The experimental data and retrieval performance

Image set	# of images	# of target images	Target ratio	Precision	Recall
butterfly	299	99	0.33	0.6682	0.9667
mountain	156	96	0.62	0.6013	0.6014
woolwork	108	36	0.33	0.6112	0.8750
flower	185	98	0.53	0.8382	0.7742
sail boat	160	111	0.69	0.6615	0.6757
waterwave	254	70	0.28	0.8291	0.4782

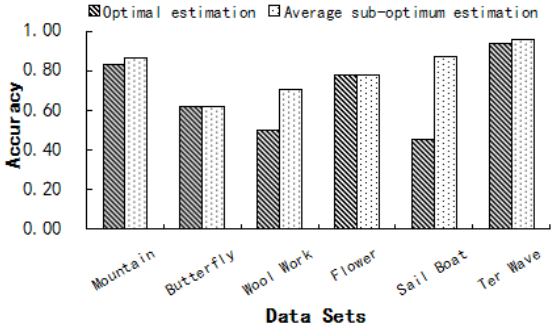

Fig. 1 The accuracy comparison of adaptive-weight preference estimation with 15% noise

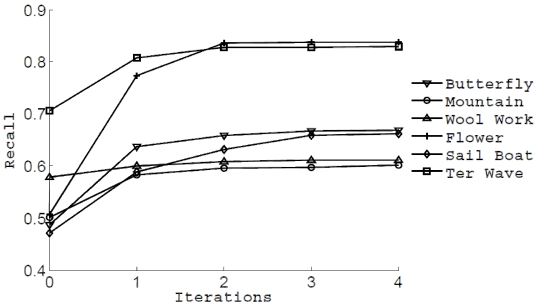

Fig. 2 The precision and convergence curves

Fig. 3 Performance comparison of user engagements

4. CONCLUSION

An adaptive-weight hybrid relevance feedback (AHRF) method for content-based image retrieval has been proposed in this paper. Taking advantage of reduced explicit grading effort and implicit browsing operations from the users, AHRF can adaptively estimate the preference degree of the browsed images and build a preference classifier for each user. The proposed method is evaluated on a subset of Corel database. The experimental results demonstrate that our method achieves good retrieval accuracy in very few iterations, providing the potential to be further used in refining image retrieval results. Future work focuses on testing and improving our method on large scale image set, multi-users conditions and real retrieval problem.

5. REFERENCES

[1] Tseng, V.S., Su, J.H., Huang, J.H. and Chen, C.J., "Integrated Mining of Visual Features, Speech Features, and Frequent Patterns for Semantic Video Annotation," *Multimedia, IEEE Transactions on* , vol.10, no.2, pp.260-267, Feb. 2008.

[2] Tseng, V. S., Su, J.H., Wang, B.W. and Lin, Y.M., "Web image annotation by fusing visual features and textual information". *Proceedings of the 2007 ACM symposium on Applied computing*, ACM, pp.1056-1060, 2007.

[3] Zhou, X.S., Huang, T.S., "Relevance feedback in image retrieval: A comprehensive review". *Multimedia systems*, Springer. vol.8, no.6, pp. 536-544, 2003.

[4] Su, J.H., Huang, W.J., Yu, P.S. and Tseng, V.S. "Efficient relevance feedback for content-based image retrieval by mining user navigation patterns", Knowledge and Data Engineering IEEE Transactions on, vol.23, no.3, pp.360-372, March 2011.

[5] Kelly, D., Belkin, N.J., "Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback". *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp.408-409, 2001.

[6] Auer, P., Hussain, Z., Kaski, S., Klami, A., Kujala, J., Laaksonen, J., Leung, A., Pasupa, K. and Shawe-Taylor, J., "Pinview: Implicit Feedback in Content-Based Image Retrieval", *Proceedings of Workshop on Applications of Pattern Analysis JMLR Workshop and Conference Proceedings*, JMLR, pp. 51-57, 2010.

[7] Faro, A., Giordano, D., Pino, C. and Spampinato, C., "Visual attention for implicit relevance feedback in a content based image retrieval", *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, pp.73--76, 2010.

[8] Nichols, D.M., "Implicit ratings and filtering", *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering*, Hungary, pp.31-36, 1997.

[9] Zhang, Y., Mo, Z., Li W., Zhao, T., "A user-driven model for Content-based image retrieval", *Proceedings APSIPA Annual Summit and Conference*, Hollywood, USA, December 2012.

[10] Grigorova, A., De Natale, F.G.B., Dagli, C. and Huang, T.S., "Content-Based Image Retrieval by Feature Adaptation and Relevance Feedback," *Multimedia, IEEE Transactions on*, vol.9, no.6, pp.1183-1192, Oct. 2007