Relevance Feedback for Spine X-ray Retrieval

Xiaoqian Xu, D.J. Lee,
Dept. of ECEn,
Brigham Young University,
Provo, UT
xiaoqian@et.byu,edu,
djlee@ee.byu.edu

S. Antani, L.R. Long
Communication Engineering Branch,
National Library of Medicine,
Bethesda, MD
antani@nlm.nih.gov,
long@nlm.nih.gov

Abstract

Relevance feedback (RF) has been an active research area in Content-based Image Retrieval (CBIR). RF intends to bridge the gap between the low-level image features and the high-level human visual perception by analyzing and employing the feedback information provided by the user. This gap becomes more evident and important in medical image retrieval due to the two distinct facts with regard to medical images: (1) subtle differences between images, even between pathological and non-pathological images; (2) subjective and different diagnosis even among experts. This paper describes a novel linear weight-updating approach for RF applying to spine x-ray image retrieval. The algorithm utilizes both positive and negative examples to gain feedback from the user. Experimental results show that the proposed approach can substantially improve the retrieval performance to better satisfy the individual user's preferences.

1. Introduction

CBIR has aroused research effort to develop efficient image features in terms of color, shape, and texture for decades. However, the gap between the low-level image features and the high-level concepts always exists and causes a bottleneck for the performance of most image features used in CBIR systems. Furthermore, mere low-level image features may not satisfy the individual user's needs considering the subjective nature of human visual perception. Since the mid-90s in the last century, RF has been proposed in aim to address this issue. The fundamental concept of RF is to establish the interaction between the user and the retrieval system and to refine the retrieval results based on the feedback provided by the user. Thus, the core of RF is a machine learning algorithm. In the literature, Neural Network (NN) and statistical approaches contribute to the main categories of RF. NN approaches require an appropriate training set and RF occurs during the training process [1-3]. NN approaches are not suitable for instant refinement of the retrieval results. Instead, in most statistical approaches RF occurs after the user is not satisfied with the retrieval results and intends to refine them. A display strategy decides what images should be provided to the user for relevance feedback. In a rather large amount literature, the retrieval results, i.e. the most similar images to the user's query, were simply displayed for the user to provide the feedback [4-6, 10]. However, this scheme could possibly get into a situation of no further improvement by over-learning and also ignore the possible useful feedback on negative images, i.e. the non-similar images. Maximum Entropy display appeared in the recent literature [7-8], which was claimed to maximize the information possibly elicited from the user.

There are various statistical approaches for machine learning. Rui et. al. proposed a straightforward and effective hierarchical weight-updating method [4]. However, there are some evident deficiencies with this method. Detailed discussion will be presented in Section 3. Probability-estimating approach also gained large research interest [7-8, 10]. Bayesian rule was used to estimate the probability of each image being the user's query in [8]. The probability was conditional given all the feedback history from the user and was updated globally in each feedback iteration. The system was built sophisticatedly while the updating process was computationally extensive especially when the size of the image database grows big. In [7], Expectation Maximization (EM) was employed to estimate the parameters, i.e. the mean and variance of the user's target distribution given a Gaussian distribution assumption. A maximum likelihood function was chosen for applying EM to make most images appear in the medium likelihood area. This method was only tested and compared with Rui's method on synthetic data. An integrated probability function for calculating the similarity between images was introduced in [10]. It was based on a posterior probability estimator and a weight updating scheme. RF from the user was used to update the weight and re-estimate the posterior probability. The method was tested on trademark images and demonstrated a retrieval performance improvement from 75% to 95% after two iterations of RF.

Few existing RF methods have been applied to medical image database. However, in aim to provide subjectively accurate medical image retrieval information online, RF becomes a necessary and indispensable part in an online medical image retrieval system. As ongoing research for spine x-ray image retrieval at the National Library of Medicine (NLM), an efficient and effective RF algorithm is needed to provide an interactive online medical information retrieval system on spine x-ray images. In this paper, a novel linear weight-updating approach is proposed and applied to spine x-ray image retrieval.

This paper is organized as follows. The retrieval approach for spine x-ray images is briefly reviewed in Section 2. The linear weight-updating approach is proposed in Section 3. In Section 4, the performance of the proposed RF approach is presented and discussed. We conclude with Section 5.

2. Shape-based retrieval approach for spine x-ray images

Due to the low contrast and low quality of x-ray images, only spine shape provides meaningful information for pathology. Our previous research work focused on developing efficient shape representations for spine x-rays retrieval [11-13]. Whole shape matching and Partial Shape Matching (PSM) have both been explored for this task. PSM enables the user to query on a specific region (possibly a pathological region) and demonstrated high retrieval accuracy in terms of pathological information [12]. However, mere PSM may not satisfy all the users' preferences. Combination of both whole shape matching and PSM according to the user's preference is an appropriate solution. From the view of the radiologist, 9-point model for spine shape was used to diagnose the osteophyte pathology. An algorithm has been developed to localize 9 points automatically according to their semantic meanings provided by the radiologist [14]. Procrustes Distance between two sets of 9 points provided a similarity measurement between two shapes. Thus, the three methods: whole shape matching, PSM, and 9 points Procrustes Distance were employed and combined to provide the final image retrieval results. Specifically, whole shape properties including compactness, elongation, and moments were used to provide whole shape matching results. Multiple open triangles with Dynamic Programming served as the shape representation to provide PSM results [12]. The merging cost, length distance, and angle distance were the components of the PSM method.

Weights exist in both levels, component level and method level. As mentioned in [4], necessary normalization ensured the meaning of adjusting the weights for emphasis on different methods or components. And there existed both intra-normalization, i.e. between components and inter-normalization, i.e. between methods. For our spine x-ray retrieval approach, the components of both PSM and whole shape matching already have the same dynamic range, respectively [12] and thus do not require intra-normalization. The similarity ranges of the three methods are the same and simply range from 0 to 1. Thus the inter-normalization is not required either.

3. Linear weight-updating approach for RF

3.1. Rui's weight-updating method

There were weights associated with each level, with the top level consisting of visual features, color, shape, and texture, the second level of a set of representations for a given feature, and the third level of possible multiple components for each representation. Gaussian model was assumed to perform the normalization. Thus all the similarity values of the feature or component under consideration need to be presented so as to calculate the mean and the standard deviation. For example, the similarity between any pair of images in the database needed to be computed for the intra-normalization. Thus this normalization process is not practical for PSM given the fact that there are infinite possible partial queries specified by the user. In addition, the similarity of each component of PSM always depends on the specific partial query and can not be presented ahead of the time. However, as mentioned in Section 2, no normalization is needed in our current spine shape retrieval system. Yet since Rui's method needed the means and deviations to update the weights, we did not implement their method for comparison. However, the discussions regarding Rui's method are detailed here.

Rui's method simply displays the most similar N_{RT} objects to the user for feedback. The user groups the N_{RT} objects into 5 categories according to his/her perception. Two different weight-updating approaches are taken for the feature level and the component level respectively. For the component level, the weights are updated as the reciprocal of the standard deviation of the component value sequence from the relevant set specified by the user. This strategy sounds reasonable on the surface. However, the measure of whether a component is a good indicator of the user's perception should be if this component is a competent to differentiate the relevant sets from the irrelevant sets. It is very likely that one component, which is able to differentiate the relevant sets from the irrelevant sets, has a larger deviation within the relevant sets than another component, which is not able to differentiate the relevant sets from the irrelevant sets. Furthermore, the weights for the feature level and the component level are updated independently and simultaneously. This causes inefficiency and possibly problems. Suppose a given feature does not perform well according to the user's feedback and thus is assigned a lower weight during the weight-updating process. In the meantime, the weights associated with the components of this feature are also updated, which actually results in making the feature a good indicator of the user's perception. Thus the feature deserves a higher weight. By addressing and solving the problems of Rui's method, a novel linear weight-updating approach is proposed in 3.3.

3.2. Retrieval system and display strategy

The retrieval system has two modes: Mode R and Mode F. Mode R displays the most similar images to the user's query while Mode F displays the images to the user for feedback. The default mode is Mode R. If the user is not satisfied with the retrieval results, the system switches to Mode F and waits for the feedback. Then the system goes back to Mode R after refining the retrieval results using the feedback information from the user. Two separate modes avoid the possible confusion and frustration from the user feeling that the system cannot present the most relevant results in the retrieval process [7].

The user can specify how many matches he/she wants to retrieve. However, for our application and experimental demonstration, Mode R retrieves and displays top 20 matches to a specific query. Mode F, however, displays both positive and negative images, i.e. 15 positive images and 5 negative images. The inclusion of both the positive images and the negative images ensures possible corrections on severe miss for the positive matches while still retains proficient feedback information on the positive matches. Specifically, the 15 positive images are randomly selected from the top 20 matches while the 5 negative images are randomly selected from the matches after the 200 most similar images.

3.3. Linear weight-updating approach

The form of relevance feedback should be as simple as possible since the user can be very ignorant about CBIR concept and all the user knows is what he/she is looking for visually. In Mode F, the user simply groups all the images displayed into three sets: relevant, not-sure, and irrelevant.

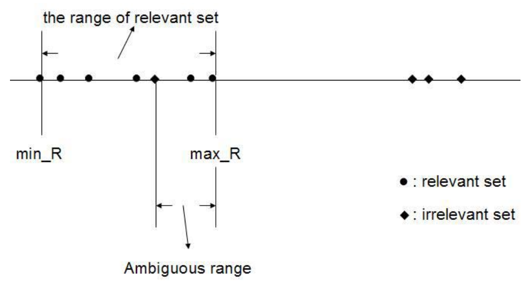

Figure 1. The distribution of the dissimilarity values of a given feature

The weights of each level, i.e. the method level and the component level, will be updated according to the same weight-updating scheme since there are no distinct reasons why they should be treated differently. Thus, in this paper, for the simplicity, the weight-updating scheme will be introduced in the component level. The dissimilarity values of a given method for all the three sets can possibly distribute as in Fig. 1. The ideal case is that no overlapping between the relevant set and the irrelevant set, which means that this method reflects the user's preference perfectly. However, the real case is usually not ideal. For all the images in the relevant set, a range of the dissimilarity values can be obtained and denoted as $[\min\_R, \max\_R]$. Very likely the dissimilarity value (denoted as d_{IR}) of an image from the irrelevant set lies in $[\min\_R, \max\_R]$. Thus an ambiguous range occurs as shown in Fig. 1. A difference related to this ambiguous range is calculated as:

$$Dif = d_{IR} - \max\_R \tag{1}$$

However, generically for every image in the irrelevant set, a difference value can be obtained through Equation (1) and the following relations hold:

$$\begin{aligned} &Dif > 0, \ \textit{no ambiguous range occurs} \\ &Dif < 0, \ \textit{ambiguous range occurs} \end{aligned} \tag{2}$$

Suppose the weights associated with all the components of this method are $W_i, i = 1,...,N$. For every occurrence of the ambiguous range in the method level, all the weights, i.e. $W_i, i = 1,...,N$, are updated once according to the linear equation (3):

$$\begin{aligned} W_i = W_i + &\max_{\substack{j \\ Dif_i - Dif_j > 0}} \left\{ \frac{Dif_i - Dif_j}{\max_{i,j}\{|Dif_i|, |Dif_j|\}} \times \frac{|W_i - W_j|}{\max\{W_i, \frac{W_i + W_j}{2}\}} \right\} \\ + &\min_{\substack{j \\ Dif_i - Dif_j < 0}} \left\{ \frac{Dif_i - Dif_j}{\max_{i,j}\{|Dif_i|, |Dif_j|\}} \times \frac{|W_i - W_j|}{\max\{W_i, \frac{W_i + W_j}{2}\}} \right\} \end{aligned} \tag{3}$$

, where Dif_i is calculated the same way as in Equation (1) but with the distribution of the dissimilarity values of the i^{th} component. This linear approach updates the weights 'dependently' by comparing the ambiguous ranges of different components, i.e. $Dif_i - Dif_j$ and their corresponding current weights, i.e. $|W_i - W_j|$. This 'dependent' updating approach ensures the true meaning of the weights since a mere weight value does not provide any information on the emphasis unless it is compared with other weights. The maximum operation on the denominators provides a gradual increase or decrease in the change of the weights and avoids big fluctuations. For $Dif_i - Dif_j > 0$, the maximum value is chosen to add to W_i, which covers all the positive updates to W_i; for $Dif_i - Dif_j < 0$, the minimum value is chosen to add to W_i, which covers all the negative updates to W_i. Thus, the maximum or minimum value update ensures the accurate update direction and proficient amount while avoids big fluctuations compared with the accumulation value update. The weights associated with the methods are updated using the same scheme by observing every occurrence of the ambiguous range in the top dissimilarity level.

Bottom-up method is employed during the weight-updating procedure as shown in Fig. 2. Specifically, the weights of the components are updated first and the dissimilarity values of the methods are updated by using the new weights. At last, the weights of the methods are updated by using the new dissimilarity values of the methods. This bottom-up method addresses the disadvantages of Rui's method by updating the weights of all levels independently and simultaneously as discussed in Section 3.1.

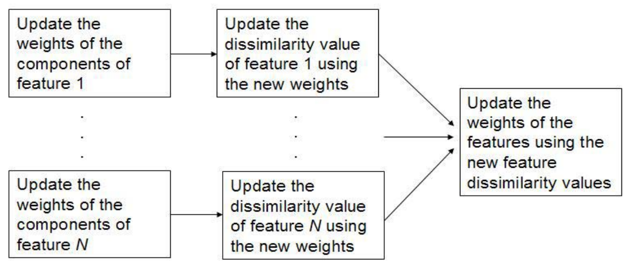

Figure 2. Bottom-up weight-updating procedure

4. Experimental results

Nearly 2,000 spine shapes segmented from 400 x-ray images at NLM were chosen to demonstrate the performance of the proposed RF algorithm. Eight queries were selected to have the user perform the relevance feedback. Fig. 3 shows the retrieval results of one of the 10 selected queries without feedback and after the first feedback iteration. Statistically, the retrieval accuracy, i.e. the number of the relevant retrieved images over the number of the total 20 retrieved images, was calculated for all the eight selected queries before and after each feedback refinement of all the first three feedback iterations. As shown in Table 1, the accuracy increased as more feedback iterations involved. Averagely, the accuracy increased from 73.75% to 80% after the first iteration of relevance feedback. After the three iterations of the relevance feedback, the accuracy reached 88.13%, which almost increased by 15% compared with the accuracy without any feedback. This RF algorithm is very efficient. This linear updating approach took less than one second to finish the

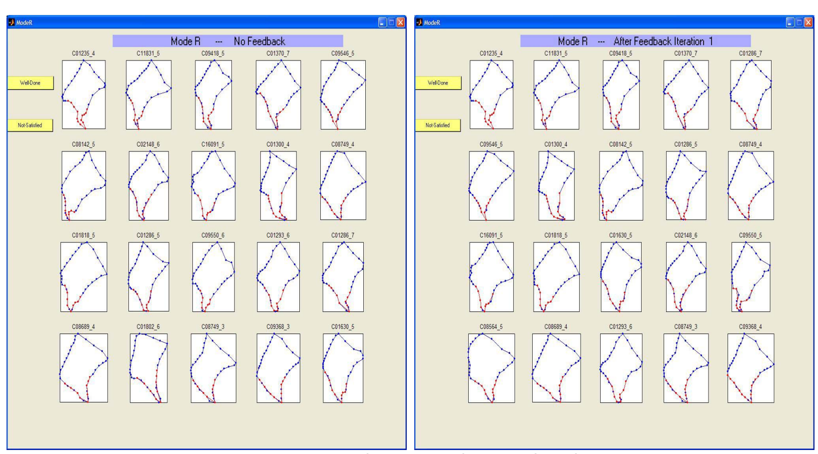

Figure 3. Retrieval results before and after the first feedback iteration

refinement after inputting the relevance feedback information.

Table 1. The retrieval accuracy of eight selected queries among top 20 matches

	Query1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Average
No feedback	70%	50%	80%	90%	45%	95%	60%	100%	73.75%
feedback 1	80%	55%	85%	100%	50%	100%	70%	Done	80%
Feedback 2	85%	65%	90%	Done	60%	Done	85%	Done	85.63%
Feedback 3	90%	75%	90%	Done	65%	Done	85%	Done	88.13%

5. Conclusions

A linear weight-updating approach for RF has been proposed in this paper. The algorithm was tested on nearly 2,000 spine x-ray shapes and performed very effectively and efficiently. It is a very promising algorithm for online medical image retrieval support. Further testing on a variety of medical images will be helpful. Also, the inclusion of other features besides shape, such as color and texture in the retrieval system will also help test the completeness of this RF algorithm.

6. References

[1] Guo-Dong Guo, Jain, A.K., Wei-Ying Ma and Hong-Jiang Zhang, "Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback", IEEE Transactions on Neural Networks, vol. 13, Issue 4, July 2002, pp. 811 – 820.

[2] Muneesawang, P. and Guan, L., "A Neural Network Approach for Learning Image Similarity in Adaptive CBIR", IEEE Fourth Workshop on Multimedia Signal Processing, Oct. 3-5, 2001, pp. 257 – 262.

[3] Breen, C., Khan, L. and Ponnusamy, A., "Image Classification Using Neural Networks and Ontologies", in The 13th International Workshop on Database and Expert Systems Applications, Sept. 2-6, 2002, pp. 98 – 102.

[4] Yong Rui, Huang, T.S., Ortega, M. and Mehrotra, S., "Relevance Feedback: A Power Tool for Interactive Content-based Image Retrieval", IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, Issue 5, Sept. 1998, pp. 644 – 655.

[5] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos, "Pichunter: Bayesian Relevance Feedback for Image Retrieval", International Conference on Pattern Recognition, vol. 3, 1996, pp. 362 – 369.

[6] S. Ghebreab, C. Jaffe and A. W.M. Smeulders, "Concept-based Retrieval of Biomedical Images", in Proceedings of SPIE Medical Imaging 2003: PACS and Integrated Medical Information Systems, vol. 5033, Feb. 18 – 20, 2003, pp. 97 – 108.

[7] Sia, K.C. and King, I., "Relevance Feedback Based on Parameter Estimation of Target Distribution", in International Joint Conference on Neural Networks, vol. 2, May 12 – 17, 2002, pp.1974 -1979.

[8] Cox, I.J., Miller, M.L., Minka, T.P., Papathomas, T.V. and Yianilos, P.N., "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments", IEEE Transactions on Image Processing, vol. 9, issue 1, Jan. 2000, pp. 20 – 37.

[9] Kyoung-Mi Lee and W. N. Street, "Incremental Feature Weight Learning and its Application to a Shape-based Query System", Pattern Recognition Letters, vol. 23, No. 7, May 2002, pp. 865 – 874.

[10] I. King and Zhong Jin, "Integrated Probability Function and its Application to Content-based Image Retrieval by Relevance Feedback", Pattern Recognition, vol. 36, No. 9, Sep. 2003, pp. 2177 – 2186.

[11] S. Antani, D.J. Lee, L.R. Long, G.R. Thoma, "Evaluation of Shape Similarity Measurement Methods for Spine X-ray Images", Journal of Visual Communication and Image Representation, Special issue: Multimedia Database Management Systems, vol. 15, No 3, 2004, pp 285-302.

[12] Xiaoqian Xu, D.J. Lee, S. Antani, L.R. Long, "Partial Shape Matching of Spine X-ray Shapes Using Dynamic Programming", in Proc. 17th IEEE Symposium on Computer-Based Medical Systems, June 2004, pp. 97 – 102.

[13] D.J. Lee, S. Antani, and L.R. Long, "Similarity Measurement Using Polygon Curve Representation and Fourier Descriptors for Shape-based Vertebral Image Retrieval", in Proceedings of SPIE Medical Imaging: Image Processing, vol. 5032, February 15-20, 2003, pp. 1283-91.

[14] X. Xu, D.J. Lee, S. Antani, and L.R. Long, "Localizing Contour Points for Indexing an X-ray Image Retrieval System", in Proceedings of The 16th IEEE Symposium on Computer-Based Medical Systems, June 26-27, 2003, pp. 169-174.