FUZZY RELEVANCE FEEDBACK IN CONTENT-BASED IMAGE RETRIEVAL SYSTEMS USING RADIAL BASIS FUNCTION NETWORK

Kim-Hui Yap and Kui Wu
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
{ekhyap, pg01537831}@ntu.edu.sg

ABSTRACT

This paper presents a new framework called fuzzy relevance feedback in interactive content-based image retrieval (CBIR) systems based on soft-decision. An efficient learning approach is proposed using a fuzzy radial basis function network (FRBFN). Conventional binary labeling schemes require a crisp decision to be made on the relevance of the retrieved images. However, user interpretation varies with respect to different information needs and perceptual subjectivity. In addition, users tend to learn from the retrieval results to further refine their information priority. Therefore, fuzzy relevance feedback is introduced in this paper to integrate the users' fuzzy interpretation of visual content into the notion of relevance feedback. Based on the users' feedbacks, an FRBFN is constructed, and the underlying parameters and network structure are optimized using a gradient-descent training strategy. Experimental results using a database of 10,000 images demonstrate the effectiveness of the proposed method.

1. INTRODUCTION

Content-based image retrieval is a process of retrieving a set of desired images from an image collection on the basis of visual contents such as color, texture, shape or spatial relationship that are present in the images. These low-level features, however, may not correspond to the users' dynamic and subjective interpretation of image contents under various circumstances. In view of this, relevance feedback, as an interactive mechanism, has been introduced to facilitate image retrieval [1-2].

User interpretation and understanding of images tend to be uncertain or imprecise due to perceptual subjectivity and changing information need under different circumstances. In this paper, we propose the notion of fuzzy relevance feedback to model the users' fuzzy interpretation of image similarity during relevance judgment process in interactive image retrieval systems, thus aiding the understanding and expression of user information need. It allows soft decisions to be made with respect to the retrieval results. Simplicity and accuracy are also achieved for the user to interact with the retrieval system.

Many relevance feedback algorithms such as query refinement [1], re-weighting [2], Bayesian learning [3], optimal learning over heuristic-based feature weighting [4-5], artificial neural networks [6], discriminant-EM algorithm [7], and kernel-based learning [8], etc., have been adopted in CBIR systems and demonstrated considerable performance improvement. In [9], an adaptive radial basis function network (ARBFN) model has been proposed for interactive image retrieval. It characterizes the query by multiple-class models associated with the relevant (positive) samples. The irrelevant (negative) samples are used to modify the models such that the models will be moved slightly away from the irrelevant samples. However, the method is heuristic as there is no adequate learning process to optimize the underling network parameters since a single-pass strategy is adopted.

In this paper, we propose an FRBFN-based method to model and learn the users' perception of image similarity in interactive image retrieval. During each feedback iteration, the relevance of the fuzzy feedbacks is evaluated using an *a posteriori* probability estimator. The network parameters undergo a supervised gradient-descent-based learning procedure by minimizing a properly chosen cost function. The trained FRBFN is then used in the next sessions to retrieve the images.

2. FUZZY RELEVANCE FEEDBACK

Traditionally, users are restricted to binary classification as to determine whether an image is "fully relevant" or "totally irrelevant" [1,3]. This process is also known as binary labeling. However, binary labeling is based on hard-decision on whether the retrieval results satisfy the user information need. It does not reflect the nature of user interpretation and understanding of images, which tends to be uncertain or imprecise due to perceptual subjectivity and changing information needs. On the other hand, multi-level labeling categorizes the positive

0-7803-9332-5/05/$20.00 ©2005 IEEE

and negative examples into several discrete levels of (ir)relevance [2,10]. The technique, however, is both inconvenient as well as imprecise as the users need to classify an image into one of the multiple levels. This conflicts with the uncertainty embedded in human perception. Users are more inclined towards using linguistic expressions such as "this image is more or less relevant" or "this image is more relevant than that one". Taken into account this problem, a fuzzy relevance feedback concept which integrates the fuzzy interpretation into the notion of relevance feedback is proposed. A fuzzy option between "relevant" and "irrelevant" is incorporated into the relevance judgment to simulate the users' decision-making process in image retrieval. Users are then allowed to provide a vague or natural description of the retrieval results in the form of fuzzy feedbacks where three choices are provided: relevant, irrelevant or fuzzy. Thus the proposed scheme reconciles the dilemma of binary or multi-level labeling by employing soft-decision instead of hard-decision on the retrieval results. A corresponding FRBFN-based learning process is developed to learn the different degrees of relevance embedded in the users' interpretation of the visual contents.

3. FRBFN LEARNING

3.1. FRBFN creation

The architecture of FRBFN is given in Figure 1. It has a structure consisting of an input layer, a Gaussian kernel layer and an output layer. The input data to FRBFN is *P*-dimensional feature vectors. They are connected to the Gaussian kernel layer which is constructed from the relevant, irrelevant, and fuzzy samples. The output layer consists of a single unit whose output value is the linear combination of all the responses from each Gaussian RBF unit.

The FRBFN is constructed dynamically based on all the accumulated training samples over previous feedback sessions. Taking into account online learning process where the users interact with the retrieval system in real time, and the training samples increase with time, we use an efficient hierarchical clustering approach to choose the FRBF centers [11]. The procedure is to cluster the samples in each relevant, irrelevant and fuzzy category, respectively. After clustering, a set of clusters is obtained and the initial FRBF center estimates are determined as the centroid of each cluster.

Let $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_K\} \subset \Re^P$ be the set of *P*-dimensional FRBF centers, and *K* be its cardinality. The Gaussian function is selected as the basis function, and the FRBFN output $F(\mathbf{x})$ for an input vector of a particular image $\mathbf{x}$, is defined as:

$$\begin{aligned} F(\mathbf{x}) &= \sum_{i=1}^{K} w_i f(\mathbf{x}, \mathbf{v}_i, \sigma_i) \\ &= \sum_{i=1}^{K} w_i \exp\left(-\frac{(\mathbf{x}-\mathbf{v}_i)^{\mathrm{T}} \Lambda (\mathbf{x}-\mathbf{v}_i)}{2\sigma_i^2} \right) \end{aligned} \quad (1)$$

where w_i is the connection weight of the output layer. $\mathbf{v}_i$ and σ_i are the center and the corresponding width of the *i*th FRBF unit. The determination of the FRBF unit width σ_i is given by:

$$\sigma_i = \gamma \cdot \min_j \left\| \mathbf{v}_i - \mathbf{v}_j \right\|, \quad j = 1,2,\ldots,K, \quad j \neq i \quad (2)$$

where γ is a factor that controls the overlapping of different FRBF units. $\Lambda = diag[\alpha_1, \ldots, \alpha_P]$ is a diagonal matrix that denotes the relative importance of different feature components, determined by the standard deviation of the positive samples.

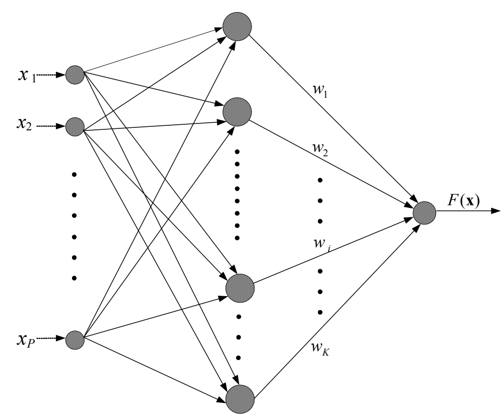

Figure 1. Architecture of FRBFN

3.2. FRBFN learning algorithm

In this study, the relevance feedback procedure is implemented as an online error correction learning by adjusting the parameters (weight, center and width) of the network. The error function is defined as:

$$E = \frac{1}{2}\sum_{j=1}^{N} e_j^2 = \frac{1}{2}\sum_{j=1}^{N}\left(Y_j - F(\mathbf{x}_j)\right)^2 \quad (3)$$

where *N* is the number of total training samples, $F(\mathbf{x}_j)$ represents the actual network output of the *j*th training sample $\mathbf{x}_j$, and Y_j is the desired network output for $\mathbf{x}_j$. We set Y_j to 1 and 0 for $\mathbf{x}_j$ associated with the positive and negative feedback, respectively. The desired output value Y_j for fuzzy feedback $\mathbf{x}_j$ is mapped to the range of [0,1]. It is estimated based on an *a posteriori* probability estimator, which combines multiple features for dealing with the uncertainty. In other words, we would like to find the probability $P(\omega_r \mid \mathbf{x}_j)$ that a fuzzy sample $\mathbf{x}_j$ belongs to the relevant class ω_r.

Let $\{\mathbf{x}_{ji}\}_{i=1}^{M}$ be the set of M features associated with a training sample $\mathbf{x}_j$, where each $\mathbf{x}_{ji}$ is a d_i- dimensional feature subvector such as color histogram, wavelet moments, among others. ω_r and ω_{ir} denote the relevant and irrelevant class respectively. The following estimation principle is used:

$$P(\omega_r \mid \mathbf{x}_j) = \frac{1}{M}\sum_{i=1}^{M} P(\omega_r \mid \mathbf{x}_{ji}) \tag{4}$$

where $P(\omega_r \mid \mathbf{x}_{ji})$ is the *a posteriori* probability for the *i*th feature vector $\mathbf{x}_{ji}$ of the fuzzy sample $\mathbf{x}_j$. According to the Bayesian theory, we have:

$$P(\omega_r \mid \mathbf{x}_{ji}) = \frac{p(\mathbf{x}_{ji} \mid \omega_r)P(\omega_r)}{p(\mathbf{x}_{ji} \mid \omega_r)P(\omega_r) + p(\mathbf{x}_{ji} \mid \omega_{ir})P(\omega_{ir})} \tag{5}$$

where $P(\omega_r), P(\omega_{ir})$ are the prior probabilities of the relevant and irrelevant classes, which can be estimated from the feedback samples. $p(\mathbf{x}_{ji} \mid \omega_r)$ and $p(\mathbf{x}_{ji} \mid \omega_{ir})$ are the class conditional probability density functions of $\mathbf{x}_{ji}$ for the relevant and irrelevant classes, respectively. Each feature vector of the relevant and irrelevant classes is assumed to be Gaussian-distributed. The probability density function is then given by:

$$p(\mathbf{x}_{ji} \mid \omega_m) = \frac{1}{(2\pi)^{d_i/2}\left|\Sigma_i^m\right|^{1/2}} \cdot \exp\left[-\frac{1}{2}(\mathbf{x}_{ji} - \mu_i^m)^{\mathrm{T}} \Sigma_i^{m\,-1}(\mathbf{x}_{ji} - \mu_i^m)\right] \tag{6}$$

where $\omega_m \in \{\omega_r, \omega_{ir}\}$. μ_i^m is the mean vector and Σ_i^m is the covariance matrix for the *i*th feature vector of class ω_m. They can be estimated using N_m training samples $\mathbf{x}_{ji}^m$ in each class:

$$\mu_i^m = \frac{1}{N_m}\sum_{j=1}^{N_m} \mathbf{x}_{ji}^m \tag{7}$$

$$\Sigma_i^m = \frac{1}{(N_m - 1)}\sum_{j=1}^{N_m} (\mathbf{x}_{ji}^m - \mu_i^m)(\mathbf{x}_{ji}^m - \mu_i^m)^{\mathrm{T}} \tag{8}$$

By minimizing the cost function E using gradient-descent method, we can update the parameters of the FRBFN:

$$\{w_i, \mathbf{v}_i, \sigma_i \mid i = 1,2,\ldots,K\} = \arg\min(E) \tag{9}$$

The weight, center and width for the *i*th RBF unit are updated as follows:

$$w_i(t+1) = w_i(t) - \eta_1 \frac{\partial E(t)}{\partial w_i(t)}, \quad i = 1,2,\ldots,K \tag{10}$$

$$\frac{\partial E(t)}{\partial w_i(t)} = -\sum_{j=1}^{N} e_j(t) f(\mathbf{x}_j, \mathbf{v}_i(t), \sigma_i(t)) \tag{11}$$

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) - \eta_2 \frac{\partial E(t)}{\partial \mathbf{v}_i(t)}, \quad i = 1,2,\ldots,K \tag{12}$$

$$\frac{\partial E(t)}{\partial \mathbf{v}_i(t)} = -w_i(t)\sum_{j=1}^{N} e_j(t) f(\mathbf{x}_j, \mathbf{v}_i(t), \sigma_i(t)) \frac{\Lambda(\mathbf{x}_j - \mathbf{v}_i(t))}{\sigma_i^2(t)} \tag{13}$$

$$\sigma_i(t+1) = \sigma_i(t) - \eta_3 \frac{\partial E(t)}{\partial \sigma_i(t)}, \quad i = 1,2,\ldots,K \tag{14}$$

$$\frac{\partial E(t)}{\partial \sigma_i(t)} = -w_i(t)\sum_{j=1}^{N} e_j(t) f(\mathbf{x}_j, \mathbf{v}_i(t), \sigma_i(t)) \frac{(\mathbf{x}_j - \mathbf{v}_i)^{\mathrm{T}} \Lambda(\mathbf{x}_j - \mathbf{v}_i)}{\sigma_i^3(t)} \tag{15}$$

where the term $e_j(t)$ is the error signal for the training sample $\mathbf{x}_j$ at the *t*th learning iteration. η_1, η_2, and η_3 are different learning parameters for w_i, $\mathbf{v}_i$, and σ_i, respectively. This parameter updating process repeats until convergence or a maximum number of iteration is reached. It is noted that the gradient-descent algorithm given in (10)-(15) does not involve back-propagation of error signal. Thus, it requires less training time to converge when compared with other neural networks such as multilayer perception.

4. EXPERIMENTAL RESULTS

The image database used in the experiment contains 10,000 color images of 100 different categories obtained from the Corel Gallery product. Color histogram, color moments and color auto-correlogram are used as the representation for color feature. Gabor wavelet and wavelet moments are used as the texture feature representation.

We perform subjective test to evaluate the effectiveness of our proposed FRBFN method. Ground truth-based objective performance measures cannot be employed because the fuzzy images selected by the users may span across different categories. Six users are invited to test the retrieval system. A total of 180 query images are used for evaluation. For each query, the top 25 retrieved images are displayed for feedback. We define the following performance measure, total retrieval accuracy (TRA), and relevant retrieval accuracy (RRA):

$$\text{TRA}=\frac{\text{relevant and fuzzy images retrieved in top } T \text{ returns}}{T} \tag{16}$$

$$\text{RRA}=\frac{\text{relevant images retrieved in top } T \text{ returns}}{T} \tag{17}$$

where T is the number of retrieved images. Since fuzzy samples satisfy the user information need up to a certain extent, they are also part of the users' desired images. Therefore, TRA is introduced to incorporate the users' fuzzy feedbacks. TRA and RRA can be considered as the upper and lower bounds of effective retrieval accuracy. Together, they give an overall idea on the performance of the FRBFN method.

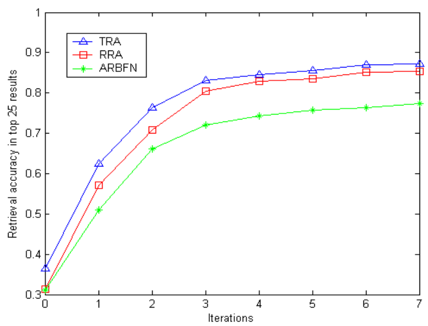

Figure 2. Performance comparison of FRBFN and ARBFN

A comparison of the retrieval performance using our FRBFN method and the ARBFN method in [9] is given in Figure 2. The zeroth iteration denotes the users' relevance judgment on the initial retrieved images based on k-nearest neighbors (K-NN) search. The retrieval performance is averaged over all queries and users. We observe that our method consistently achieves a higher retrieval accuracy than the ARBFN method. Further it is observed that to achieve a specific retrieval accuracy, the FRBFN method requires less number of iterations when compared to the ARBFN method. Coupled with fuzzy relevance feedback, the FRBFN method is more adaptive towards the user information need. All users rank the FRBFN method higher than the ARBFN method in terms of capturing their perceptual subjectivity and information needs.

5. CONCLUSION

This paper presents a new notion of fuzzy relevance feedback and the corresponding FRBFN-based framework for integrating the users' imprecise interpretation of image similarity into interactive CBIR systems. In contrast to conventional relevance feedback approaches that are based on binary or multi-level labeling, our method provides a natural and flexible way to express the users' preferences. Experimental results demonstrate that our FRBFN approach is superior in retrieval performance, and is more effective in addressing different users' information needs.

6. REFERENCES

[1] Y. Rui, T.S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," *Proc. IEEE Int. Conf. on Image Processing*, Washington D.C., USA, pp. 815-818, 1997.

[2] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. on Circuits and Video Technology*, vol. 8, no. 5, pp. 644-655, 1998.

[3] N. Vasconcelos and A. Lippman, "Learning from user feedback in image retrieval systems," *Proc of NIPS'99*, Denver, Colorado, 1999.

[4] Y. Ishikawa and R. Subramanya, "MindReader: Query database through multiple examples," *Proc. of Int. Conf. on Very Large Data Bases*, New York, USA, 1998.

[5] Y. Rui and T.S. Huang, "Optimizing learning in image retrieval," *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol.1, pp. 236-243, June 2000.

[6] J. Laaksonen, M. Koskela, and E. Oja, "PicSom-self-organizing image retrieval with MPEG-7 content descriptions," *IEEE Trans. on Neural Network*, vol. 13, no. 4, pp. 841-853, July 2002.

[7] Y. Wu, Q. Tian, and T.S. Huang, "Discriminant-EM algorithm with application to image retrieval," *Proc of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, South Carolina, 2000.

[8] S. Tong and E. Chang, "Support vector machine active leaning for image retrieval," *Proc of 9th ACM Conference on Multimedia*, Ottawa Canada, 2001.

[9] P. Muneesawang and L. Guan, "Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture," *IEEE Trans. on Neural Networks*, vol. 13, no. 4, pp. 821-834, July 2002.

[10] X.S. Zhou and T.S. Huang, "Small sample learning during multimedia retrieval using BiasMap," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, Hawaii, Dec.2001.

[11] A.K. Jain, M.N. Murty, and P.J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, Vol. 31, No. 3, pp. 264-323, 1999.