AN EFFICIENT RADIAL BASIS FUNCTION NETWORK APPROACH FOR CONTENT-BASED IMAGE RETRIEVAL

Kui Wu and Kim-Hui Yap

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

ABSTRACT

In this paper, an efficient approach using radial basis function network (RBFN) with online learning capability is proposed for interactive content-based image retrieval (CBIR) systems. Based on the users' feedbacks, an RBFN is constructed, and the underlying parameters and network structure are adjusted adaptively using a training strategy. To capture the users' perceptual consistency in similarity, an error function is expressed in terms of accumulated training samples across all feedback sessions. Experimental results using a database of 10,000 images demonstrate the effectiveness of the proposed method.

1. INTRODUCTION

The need for efficient access to image data is growing rapidly in many fields, ranging from art galleries, digital libraries, biomedicine to military and education. Content-based image retrieval (CBIR) has been developed as an effective way of accessing image data [1-4]. It interprets user information need based on a set of low-level visual features (color, texture, shape) extracted from the images. However, these features may not correspond to user interpretation and understanding of image content. Thus, a semantic gap exists between the high-level concepts and the low-level features in CBIR. Furthermore, user interpretation depends on individual's subjectivity, and may change accordingly throughout the searching process. Relevance feedback (RF), as an interactive mechanism, has been introduced to facilitate image retrieval. The user is integrated into the retrieval systems to provide his/her evaluation on the retrieval results. The systems then learn from the feedbacks to retrieve a new set of images that better satisfy the user information need.

Many relevance feedback algorithms such as query refinement [2], re-weighting [5], Bayesian or probabilistic learning [6], optimal learning over heuristic-based feature weighting [7], artificial neural networks [8], discriminant-EM algorithm [9], and kernel-based learning [10], etc., have been adopted in CBIR systems and demonstrated considerable performance improvement. In [11], a Gaussian Mixture Model (GMM) approach has been proposed for interactive image retrieval. It characterizes the query by multiple-class models, an inherent strategy of local modeling, and associates those relevant (positive) samples as the models. The irrelevant (negative) samples are used to modify the models such that the models will be moved slightly away from the irrelevant samples. As for similarity evaluation, a radial basis function network is used, which associates each model with a Gaussian kernel that describes the distribution of input data. However, the method is heuristic as there is no learning process to optimize the underling network parameters.

In this paper, we propose an RBFN-based method to model and learn the users' perception of image similarity in interactive image retrieval. Based on both the relevant and irrelevant samples, the network parameters undergo a supervised gradient-descent-based learning procedure by minimizing a properly chosen error function. Most learning methods only acquire knowledge from the current query session. We utilize all the feedback examples accumulated from previous query sessions to adaptively adjust the network structure and parameters using an efficient training strategy, thus speeding up the learning process to achieve improved retrieval performance.

2. GAUSSIAN MIXTURE MODEL

In [11], a query transaction is characterized by multiple-class models, which are associated with the set of positive samples, $\{\mathbf{v}_m\}_{m=1}^{M}$, and M be its size. Each $\mathbf{v_m}$ is described by a Gaussian distribution:

$$G_m(\mathbf{x}, \mathbf{v}_m, \sigma_m) = \exp\left(-\frac{(\mathbf{x}-\mathbf{v}_m)^{\mathrm{T}}(\mathbf{x}-\mathbf{v}_m)}{2\sigma_m^2}\right) \tag{1}$$

where σ_m is a smoothing parameter defined as:

$$\sigma_m = \delta \cdot \min_i \left\|\mathbf{v}_m - \mathbf{v}_i\right\|, \qquad i = 1, 2, ..., M \tag{2}$$

with $\delta = 0.5$ being an overlapping factor. The negative samples $\{\mathbf{x}_n^{'}\}_{n=1}^{N}$ are used to modify the models, so as to move each $\mathbf{v}_m$ slightly away from the negative samples:

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) - \eta(t)\left[x_n^{'}(t) - \mathbf{v}_i(t)\right] \tag{3}$$

where N is the number of the negative samples. $\eta(t)$ is the learning constant which decreases with the number of iterations *t*, $0 \le \eta(t) \le 1$. The similarity evaluation is then performed based on linear combination of the output G_m:

$$F(\mathbf{x}) = \sum_{m=1}^{M} G_m(\mathbf{x}, \mathbf{v}_m, \sigma_m) \tag{4}$$

The overall output $F(\mathbf{x})$ for an input data $\mathbf{x}$ from the database functions as a similarity measure, namely, those images $\mathbf{x}$ in the database with higher values are more likely to be relevant to the user.

Although the assumption of Gaussian Mixture Model in [11] is reasonable, the location of the RBF unit centers is heuristic. Further, it places equal emphasis on every positive sample. However, different positive images may have different degrees of relevance, thus contributing unequally to the user information need. This single-pass strategy does not employ adequate training procedure to adjust the parameters. In addition, it is concerned with only short-term learning based on current feedback session.

3. RBFN-BASED LEARNING

3.1. RBFN Creation

Due to its fast learning speed, simple structure, local modeling and global generalization power, RBF network is adopted in our system to learn the input-output mapping. It is constructed dynamically based on both the positive and negative training samples, since any training sample, whether desired or undesired, contains some information provided by a user.

The architecture of RBFN is given in Fig. 1. It has a topological structure including an input layer, a Gaussian kernel layer and an output layer. The input data to RBFN is a *P*-dimensional feature vector connected to the Gaussian kernel layer which is associated with the positive and negative 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.

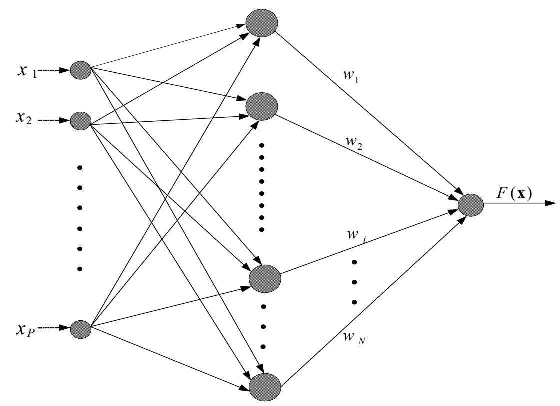

Fig. 1 Architecture of RBFN

Let $V = \{\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_K\} \subset \Re^P$ be the initial set of all *P*-dimensional training samples, which are the initial feedback samples, relevant and irrelevant, and *K* be its size. The Gaussian function is selected as the basis function, and the RBF network output $F(\mathbf{x})$ for an input vector of a particular image $\mathbf{x}$, is defined as:

$$\begin{aligned} F(\mathbf{x}) &= \sum_{i=1}^{K} w_i f(\mathbf{x}, \mathbf{v}_i, \sigma_i) \\ &= \sum_{i=1}^{K} w_i \exp\left(-\frac{(\mathbf{x}-\mathbf{v}_i)^{\mathrm{T}} \Lambda (\mathbf{x}-\mathbf{v}_i)}{2\sigma_i^2}\right) \end{aligned} \tag{5}$$

where w_i is the connection weight of the output layer. $\mathbf{v}_i, \sigma_i$ are the center and corresponding width of the *i*th RBF unit. During the initialization process, we assign a positive weight of $w_i = 1$ to each positive sample and a negative weight of $w_i = -0.5$ to each negative sample. The determination of the RBF unit width σ_i is given by:

$$\sigma_i = \gamma \cdot \min_j \left\| \mathbf{v}_i - \mathbf{v}_j \right\|, \qquad j = 1, 2, \ldots, K, \qquad j \ne i \tag{6}$$

where γ is a factor that controls the overlapping of different RBF 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.

3.2. Learning of RBFN Parameters

When a retrieved image $\mathbf{x}_j$ is marked as relevant by the user, we set the desired network output Y_j to 1. On the other hand, for an irrelevant image, the desired output is set to 0.

$$Y_j = \begin{cases} 1 & \text{if } \mathbf{x}_j \text{ is relevant} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

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

$$E = \frac{1}{2}\sum_{j=1}^{N} e_j^2 = \frac{1}{2}\sum_{j=1}^{N}\left(Y_j - F(\mathbf{x}_j)\right)^2 \quad (8)$$

where N is the number of all training samples, Y_j and $F(\mathbf{x}_j)$ represent the desired and actual output of the jth sample, respectively. By minimizing the cost function E using gradient-descent method, we can update all the parameters of the RBF neural network:

$$\mathbf{v}_i, w_i, \sigma_i = \arg\min(E) \quad (9)$$

The weight, center and width for the ith RBF unit are updated as follows:

$$w_i(t+1) = w_i(t) - \eta_1 \frac{\partial E(t)}{\partial w_i(t)} \quad (10)$$

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

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) - \eta_2 \frac{\partial E(t)}{\partial \mathbf{v}_i(t)} \quad (12)$$

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

$$\sigma_i(t+1) = \sigma_i(t) - \eta_3 \frac{\partial E(t)}{\partial \sigma_i(t)} \quad (14)$$

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

where $\eta_{1,}$ η_2, and η_3 are different learning parameters for $w_i, \mathbf{v}_i$, and σ_i, respectively.

3.3. Long-Term Learning Using Selective Training

Utilizing the information from the current query session only is a process of short-term learning, as opposed to long-term learning which acquires the knowledge over many feedback sessions. To capture a user's perceptual consistency in image similarity, we adopt the long-term learning strategy for relevance feedback. Since the training samples have some overlapping between consecutive feedback sections, re-training the whole network based on all the training samples is unnecessary and time consuming. Thus, in each feedback session, after comparing the newly retrieved images with previous samples, we insert new units into the existing RBFN that has been trained in previous iterations. The modified updating procedure is performed on the newly inserted RBF units. This selective training scheme is fast in that only a subset of the training samples is trained. In addition, it is effective since it takes into account information accumulated over different feedback sessions.

4. EXPERMENTAL RESULTS

The image database used in the experiment contains 10,000 color images of 100 different categories obtained from the Corel Gallery product (Corel, 1999). The visual features used are color- and texture-based. 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.

The performance is evaluated by selecting 100 query images, one from each category. Seven iterations of feedbacks are recorded. The following performance measure: Average Precision (APR) is adopted:

$$\overline{P_r} = \frac{1}{N}\sum_{i=1}^{N} P_i \quad (16)$$

where N is the number of selected queries, N=100 in this experiment. P_i is defined by:

$$P_i = \frac{\text{Number of relevant images retrieved, } N_r}{\text{Total number of retrieved images , } N_{RT}} \times 100\% \quad (17)$$

where $N_{RT} = 25$. A comparison of the retrieval performance using our RBFN method and the GMM-based method in [11] is given in Fig. 2.

We can observe that our RBFN method consistently achieves a higher APR than the GMM method. The APR of the RBFN method increases quickly in the initial stage, and converges to reach around 0.83 after five feedback iterations, as opposed to 0.73 offered by the GMM method. This is a desirable property, since it provides significant improvement on the retrieval results quickly using RBFN. It is observed that to achieve a specific APR, RBFN method requires less number of iterations when compared to the GMM method. This is because RBFN utilizes examples from all feedback iterations, and adaptively adjusts the parameters to achieve better performance.

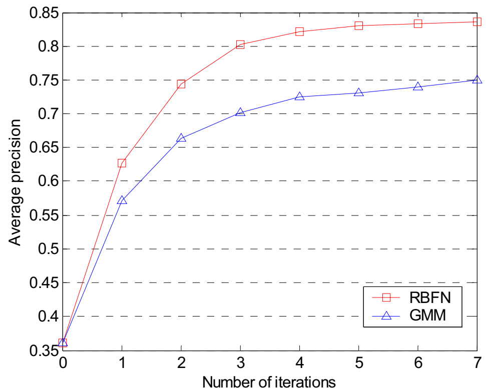

Fig. 2 Comparison of Average Precision (APR) Curve

5. CONCLUSION

This paper presents an RBFN-based relevance feedback approach for interpreting users' perception of image similarity in interactive CBIR systems. During the feedback iterations, an RBFN is constructed and progressively adjusted to obtain the proper parameters. An efficient online training algorithm is proposed to achieve fast and better retrieval results. Experimental results confirm the effectiveness of our proposed RBFN approach.

6. REFERENCES

[1] M. Flickher, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D.Steele, and P. Yanker, "Query by image and video content: The QBIC system," *IEEE Computer*, vol. 28, no. 9, pp. 23-32, Sept. 1995.

[2] 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.

[3] Amarnath Gupta, Ramesh Jain, "Visual information retrieval," *Communications of ACM*, vol. 40, no. 5, pp.70-79, May 1997.

[4] J. R. Smith, S.-F. Chang, "VisualSEEk: a fully automated content based image query system," *Proc. ACM Multimedia*, November 1996.

[5] 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.

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

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

[8] 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.

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

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

[11] 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.