GENERALIZED ADAPTIVE PERSONALIZATION OF MULTIMEDIA CONTENT

Nikolaos D. Doulamis

National Technical University of Athens, Dept. of Electrical & Computer Engineering

ABSTRACT

Modeling multimedia content by identifying semantically meaningful entities can be arduous because it is difficult to simulate human perception. However, by creating an algorithm to respond interactively to user preference, content-retrieval systems can become more efficient and easier to use. In this paper, we investigate adaptive relevance feedback algorithms for interactive multimedia content personalization. In particular two interesting scenarios are examined. The first uses a weighted cross correlation similarity measure for ranking multimedia data. The second exploits concepts of functional analysis to model the similarity measure as a non-linear function, the type of which is estimated by the users' preferences. The algorithms are computationally efficient and they can be recursively implemented.

1. INTRODUCTION

Over the past decade, the amount of existing multimedia data has exploded because of the proliferation of low-cost devices designed for capturing and encoding it. But there has also been a corresponding increase in the awareness about the challenges associated with managing that data [1]. Emerging applications designed to organize, retrieve, and abstract multimedia data require efficient technologies to be truly useful. Such applications are typically designed to help users interact more effectively with their content by addressing issues associated specifically with retrieval and personalization [2]. For example, the Moving Picture Expert Group has developed the MPEG-7 standard to provide a rich set of standardized tools for describing multimedia content [3].

However, even with such standards, managing multimedia content presents many unique challenges that we should address to make modeling and describing multimedia semantic content easier. Humans perceive and characterize content using high-level concepts—such as action, romance, comedy, or emotional factors—that are not directly related to the content's technical attributes. A piece of music, for example, can evoke many different feelings in different people. But without an interactive, adaptive system, it would be almost impossible to use human feelings as a way to store and retrieve data. An effective way to organize, retrieve, and navigate multimedia information would be to use these kinds of human perceptions to rank the data, navigate it, then update it according to the user's perceptions.

One of the interactive learning techniques that might lead to this kind of technology is called *relevance feedback*, a strategy originally developed in traditional text-based information retrieval systems. In a relevance-feedback approach, the human is part of the multimedia-management process, which means that the user evaluates the results provided by the system; the system then adapts its performance according to the user's demands and preferences. Relevance feedback, apart from accounting for the human subjectivity in perceiving the content, eliminates the gap between high-level semantics and low-level features, which are often used for content description and modeling [4].

Recently, researchers have extended relevance-feedback algorithms from text-based information retrieval to content-based image retrieval (CBIR) systems. In a relevance-feedback approach, a heuristic scheme performs the weight updating on the basis of the image variations [5]. The first approaches that used this technology have been described elsewhere [6], [7]. To avoid some of the problems associated with this approach, researchers developed a hierarchical model for decomposing the feature vectors into vectors of smaller size [8]. But even this scheme presented difficulties in cases where the CBIR systems were so large that the hierarchical model could not be applied easily [9].

This article explores relevance-feedback algorithms for interactive multimedia content personalization. In particular, it discusses two scenarios. The first relies on a weighted cross-correlation metric for the similarity measure, while the second confronts the relevance-feedback problem in the most general form by considering that the similarity measure can be almost any type of nonlinear function. In the first scenario, we propose an optimal and computationally efficient weight-updating strategy that uses the cross-correlation criterion as the similarity measure. Correlation is a normalized measure that expresses how similar two pieces of information are. It provides a similarity measure that's closer to human perception than the conventional Euclidean-distance measure found in earlier approaches. This scheme could be recursively implemented in case of multiple-feedback iterations to increase system flexibility and efficiency.

The second scenario exploits concepts derived from functional analysis [10] to estimate the similarity measure on the basis of a set of selected data that expresses the current users' needs and preferences.

2. MULTIMEDIA CONTENT MODELING

We characterize a multimedia object, in general, with a set of descriptors extracted to model its content and the similarity metric used to determine how similar or dissimilar two multimedia objects are. To provide a compact and meaningful visual-content representation, some approaches construct histograms of the extracted descriptors [11]. Usually, a content-classification scheme adopts a histogram construction by allowing each descriptor to belong only to one class [12]. Recent research, however, indicates that a better visual content representation we can achieve by allowing each descriptor to belong to several classes—with different degrees of membership. In certain cases, we adopt this technique, which results in a fuzzy organization scheme [13].

We have considered two types of descriptors: global-based and object-based [14]. The first refers to global visual characteristics, such as the global image color, texture, and motion. The second concerns features of image objects, as obtained by applying a segmentation algorithm to the image—object color, size, and location.

As mentioned above, the most commonly used similarity measure for multimedia data retrieval is the Euclidean distance, where in its generalized form is defined as

$$d(\mathbf{f}_q,\mathbf{f}_i) = (\mathbf{f}_q - \mathbf{f}_i)^T \cdot \mathbf{W} \cdot (\mathbf{f}_q - \mathbf{f}_i) \quad (1)$$

In Equation 1, vector $\mathbf{f}_q$ corresponds to the feature vector of the query image submitted by the user, while $\mathbf{f}_i$ corresponds to the feature vector of the image in the database. The $\boldsymbol{W}$ represents a real symmetric matrix, which contains the weights that regulate the degree of importance of the feature elements to the similarity measure. The Euclidean distance doesn't directly express the similarity of two feature vectors. It is sensitive to feature vector scaling and translation.

Another similarity measure is cross-correlation, which is a normalized metric that expresses how similar two feature vectors are. Cross-correlation remains unchanged with respect to feature-vector scaling and translation. In our case, we adopt a parametric weighted correlation-based similarity measure for relevance feedback [15] as in

$$\rho_{\mathbf{w}}(\mathbf{f}_q,\mathbf{f}_i) = \frac{\sum_{k=1}^{P=Q^L} w_k \cdot f_{q,k} \cdot f_{i,k}}{\sqrt{\sum_{k=1}^{P=Q^L} w_k^2 \cdot f_{q,k}^2} \cdot \sqrt{\sum_{k=1}^{P=Q^L} f_{i,k}^2}} \quad (2)$$

In Equation 2, $f_{q,k}$ and $f_{i,k}$ are the elements of vectors $\mathbf{f}_q$ and $\mathbf{f}_i$, respectively. Variable P indicates the size of feature vector $\mathbf{f}_i$, while parameters w_k indicates the relevance of the element of the query feature vector.

Although a correlation-based similarity measure can provide better characterization of multimedia content than the generalized Euclidean distance measure, its weak point is that it permits regulation only of the weighted factors w_k, assuming a constant similarity measure. We could implement a more powerful and efficient approach by permitting the similarity measure to be of any nonlinear generic type. Thus,

$$d(\mathbf{f}_q,\mathbf{f}_i) = g(\mathbf{f}_q - \mathbf{f}_i) \quad (3)$$

3. CORRELATION-BASED RELEVANCE FEEDBACK

In this section, we discuss the relevance-feedback scheme that adopts the correlation-based similarity measure shown in Equation 2.

3.1 Single relevance feedback

Single relevance feedback refers to cases in which only one interaction is adequate to adapt the system response to the current users' information needs and preferences. In these cases, the relevance information, as expressed by a set of selected relevant or irrelevant samples, is fed back to the system to update or refine the similarity measure weights $\mathbf{w}$ in Equation 2. The weights are adapted to regulate the degree of relevance of feature components to the similarity measure. In particular, the weights are updated so that after the feedback iteration the correlation of the query feature vector $\mathbf{f}_q$ and the feature vectors of all selected relevant images is maximized, whereas the correlation over all irrelevant images is minimized

$$C(\mathbf{w}) = \sum_{i=1}^{m} \rho_{\mathbf{w}}(\mathbf{f}_q,\mathbf{y}_i) = \sum_{i=1}^{m} \frac{\sum_{k=1}^{P=Q^L} w_k \cdot f_{q,k} \cdot y_{i,k} \cdot \eta_i}{\sqrt{\sum_{k=1}^{P=Q^L} w_k^2 \cdot f_{q,k}^2} \cdot \sqrt{\sum_{k=1}^{P=Q^L} y_{i,k}^2}} \quad (4)$$

In Equation 4, $\mathbf{y}_i$, with $i = 1, \ldots, m$ are the feature vectors of the images selected by the user as relevant or irrelevant to the original query, characterized by vector $\mathbf{f}_q$. The number m of selected samples is smaller than or equal to the number M of the total retrieved data. That is, ($m \le M$). Scalar η_i expresses the degree of relevance of the selected samples provided by the user.

The system obtains the optimal weights $\mathbf{w}$ by setting the derivatives of Equation 4 for all weights w_n where $n = 1, \ldots, P$ equal to zero. However, one weight is a free variable and can't be estimated by maximizing Equation 4. For this reason, the system needs an additional constraint to restrict the weight norm to a constant value. In our case, and without loss of generality, we assume that $\|\mathbf{w}\|_2 = 1$. Thus, we can create weight updating to satisfy user information needs with the following constraint maximization problem,

$$\hat{\mathbf{w}} = \arg\max_{\mathbf{w}} C(\mathbf{w}) = \sum_{i=1}^{m} \rho_{\mathbf{w}}(\mathbf{f}_q, \eta_i \cdot \mathbf{y}_i), \text{ subject to } \|\mathbf{w}\|_2 = 1 \quad (5)$$

On the basis of the constraint-maximization problem shown in Equation 5, we calculate the optimal weights $\hat{\mathbf{w}}$ to adapt the system response to the current users' information needs and preferences.

3.2 Multiple relevance feedback

In cases in which single relevance feedback isn't adequate for updating the system response to the current users' information needs and preferences, a second iteration of the weight update mechanism would be necessary. By repeating this procedure several times, we can initiate a multiple relevance feedback iteration scheme. In a multiple-feedback iteration, we apply a recursive implementation of the algorithm for estimating the optimal weights $\hat{\mathbf{w}}$.

In multiple relevance feedback, vectors $\mathbf{f}_q$ and $\mathbf{y}_i$ are discrete time sequences $\mathbf{f}_Q(r)$ and $\mathbf{y}_i(r)$, where r corresponds to the iteration index. Similarly, we assume that at each iteration, $m(r)$ images are considered as relevant or irrelevant. Then, the optimal weights $\hat{\mathbf{w}}$ at feedback iteration r are given as

$$\hat{\mathbf{w}}(r) = \arg\max_{\mathbf{w}} C(\mathbf{w}(r)) = \sum_{k=0}^{r} \lambda^{r-k} \sum_{i=1}^{m(k)} \rho_{\mathbf{w}}(\mathbf{f}_q, \eta_i \cdot \mathbf{y}_i(k))$$

$$\text{Subject to } \|\mathbf{w}(r)\|_2 = 1 \quad (6)$$

In Equation 6, λ ($0 < \lambda < 1$) is a forgetting factor that regulates the importance of the selected images at previous feedback iterations. Using the aforementioned methodology, we can derived to a recursive implementation of weights $\mathbf{w}(r)$ with respect to the previous feedback iterations.

4. GENERALIZED RELEVANCE FEEDBACK

The main difficulty of the second case dealing with generalized relevance feedback is that function $g(\cdot)$ is actually unknown.

However, using functional analysis, we can express any continuous nonlinear function $g(\cdot)$ as a parametric relation of known functional components $\Phi_l(\cdot)$ within any degree of accuracy.

The generalized distance $g(\cdot)$ is expressed as a relation of model parameters v_l, $w_{k,l}$. Parameters v_l, $w_{k,l}$ aren't related to the weighted factors. Instead, they express the coefficients on which the function $g(\cdot)$ is expanded to the respective functional components.

4.1 Optimal recursive similarity

We recursively estimate the contribution of each functional component—the parameters v_l, $w_{k,l}$—to the similarity metric through an efficient online learning strategy. In particular, we perform the adaptation so that the current selected content is trusted as much as possible without having to modify the already estimated similarity measure. The first condition means that the algorithm updates the system response to satisfy the current users' information needs and preferences as much as possible. On the other hand, the second condition implies that the adaptation should be performed so that the knowledge obtained by the previously selected samples undergoes minimal degradation. We express the first condition as

$$d^{(r+1)}(\mathbf{f}_q - \mathbf{f}_i) \approx R_i, \text{ with } i \in S^{(r)} \quad (7)$$

In Equation 7, $d^{(r+1)}(r)$ expresses the nonlinear similarity measure at the feedback iteration $(r + 1)$ of the algorithm. The R_i refers to the relevance degree of the selected images. Negative values of R_i correspond to images of irrelevant content, whereas positive values of R_i correspond to relevant images. The set $S^{(r)}$ contains all selected images at the r feedback iteration.

The main difficulty in solving Equation 7 is that function $d^{(r+1)}(r)$ is nonlinear. For this reason, we assume that a small modification of model parameters v_l, $w_{k,l}$ is adequate to satisfy Equation 7. We express this condition as

$$\mathbf{w}(r+1) = \mathbf{w}(r) + \Delta\mathbf{w} \quad (8)$$

In Equation 8, $\mathbf{w}(r) = [\cdots v_l(r) \cdots w_{k,l}(r) \cdots]^T$ refers to a vector containing all the coefficients $v_l(r)$, $w_{k,l}(r)$ at iteration r. Equation 8 means that instead of estimating the model parameters $\mathbf{w}(r + 1)$, we must only estimate the perturbation $\Delta\mathbf{w}$ to find the new model parameters.

On the basis of Equation 8, we can express the second condition, which refers to the minimal degradation of the already obtained knowledge, as

$$\text{minimize } \|\Delta\mathbf{w}\|_2 \text{ or equivalently } (\Delta\mathbf{w})^T \cdot \Delta\mathbf{w} \quad (9)$$

We perform the minimization using Lagrange multipliers and imposing e a first-order Taylor series expansion. The minimization results in a recursive estimation of the model parameters with respect to the previous feedback iteration.

5. EXPERIMENTAL RESULTS

For the purpose of our experiments, we used the image database of the National Technical University of Athens enhanced by key-frames obtained from digitalizing video sequences of the Hellenic Radio-Television broadcasting channel archives. In video sequences, we extracted key-frames using a video-summarization algorithm [14]. The overall data set consists of around 15,000 images that cover a wide variety of content. All images have been annotated by domain professionals and put into 80 categories, such as "space equipment," "tigers & lions," "fractals," and so on.

For conducting the experiments, we considered all images belonging to the same category in the database as relevant, and we considered the remaining images to be irrelevant. For visual-content representation, we extracted several descriptors and organized them according to the fuzzy-formulation scheme [13].

To evaluate the retrieval performance standard, we used quantitative measurements such as the Precision-Recall curve [16] and the Average Normalized Modified Retrieval Rank (ANMRR) [17].

We submitted around 3,000 randomly selected images to the system and we examined the average system response as expressed by the precision-recall curve and the ANMRR criterion. In these experiments, we considered all retrieved images to be relevant only if they belong to the same category as the query image. Figure 1 presents the average precision-recall curve obtained at the fifth feedback iteration for different relevance feedback algorithms. As expected, we achieved the best precision for every recall value using our second method—the generalized relevance feedback algorithm. The second best performance was our first method.

Table 1 shows the ANMRR values measured for the two proposed relevance feedback algorithms along with the results using algorithms from the other research. The generalized method provides the smallest values of the *ANMRR* measure, with the correlation-based method coming in second. This means that the presented algorithms not only yield the best retrieval results, but also provides the smallest ranking of the relevant images.

Figure 2 shows the precision values with respect to feedback iterations for 10 percent and 30 percent recall for all algorithms. We observed that the improvement ratio decreases, which means that beyond a certain point, we can only accomplish a slight increase in precision. Both our proposed schemes outperform the other three for every feedback iteration.

The correlation-based relevance feedback schemes outperform the Euclidean-based ones because correlation is a more appropriate metric for expressing the similarity of two feature vectors. Furthermore, it's robust to feature-vector scaling and translation. The most efficient relevance-feedback algorithm is the generalized one because it can adapt not only the importance of the feature vector elements but also the similarity measure type. In this way, the system can more effectively update its response to the current user's needs and preferences. Another significant advantage of our proposed algorithms is the recursive implementation in cases of multiple-feedback iterations.

6. REFERENCES

[1] M.C. Angelides, "Multimedia Information Systems," *Prospects Focus on Information Technology Magazine*, 2001, pp. 90-91.

[2] K.N. Ngan et al., "Special Issue on Segmentation, Description and Retrieval of Video Content," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 5, Sept. 1998, pp. 521-524.

[3] MPEG-7 Requirements Group, "MPEG-7: Context, Objectives, and Technical Roadmap, V.12," Vancouver, July 1999, ISO/IEC SC29/WG11 N2861.

[4] Y. Rui et al., "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval," *IEEE Trans. Circuits. Systems for Video Technology*, vol. 8, no. 5, Sept. 1998, pp. 644-655.

[5] I. Cox et al., "Pichunter: Bayesian Relevance Feedback for Image Retrieval," *Proc. Int'l Conf. Pattern Recognition*, vol. 3, 1996, pp. 362-369.

[6] A.D. Doulamis, "Interactive Content-Based Retrieval in Video Databases Using Fuzzy Classification and Relevance Feedback," *Proc. IEEE Int'l Conf. Multimedia Computing and Systems*, vol. 2, June 1999, pp. 954-958.

[7] Y. Ishikawa, R. Subramanya and C. Faloutsos, "Mindreader: Query Databases through Multiple Examples," *Proc. 24th VLDB Conf.*, New York, USA, 1998.

[8] Y. Rui and T.S. Huang, "Optimizing Learning in Image Retrieval," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, June 2000.

[9] Xiang Sean Zhou and T.S. Huang, "Small Sample Learning during Multimedia Retrieval Using BiasMap," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Dec. 2001.

[10] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1989.

[11] T. Sikora, "The MPEG-7 Visual Standard for Content Description: An Overview," *IEEE Trans. Circuits And Systems For Video Technology*, vol. 11, no. 6, June 2001, pp. 696-702.

[12] M. Flickner et al., "Query by Image and Video Content: the QBIC System," *IEEE Computer*, Sept. 1995, pp. 23-32.

[13] A. Doulamis, N. Doulamis, and S. Kollias, "A Fuzzy Video Content Representation for Video Summarization and Content-based Retrieval," *Signal Processing*, vol. 80, June 2000, pp. 1049-1067.

[14] Y. Avrithis, "Optimization Methods for Key Frames and Scenes Extraction," *J. Computer Vision and Image Understanding*, vol. 75, nos. 1/2, July/Aug. 1999, pp. 3-24.

[15] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw Hill, 1984.

[16] G. Salton and M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill, 1982.

[17] "MPEG-7 Visual Part of eXperimentation Model Version 2.0," MPEG-7 Output Document ISO/MPEG, Dec 1999.

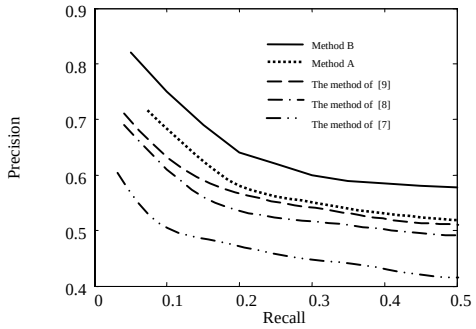

Figure 1. Relevance feedback performance of both the proposed schemes and the methods of [7], [8] and [9] as expressed by the average precision-recall curve at the 5th feedback iteration.

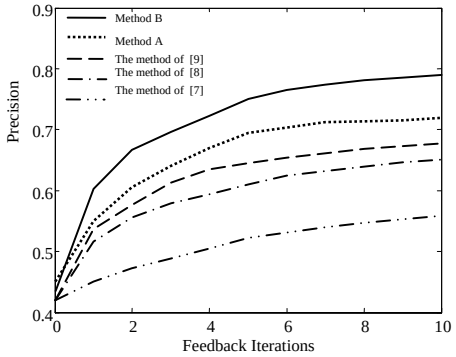

Figure 2. Precision values versus the number of feedback iterations.

Table 1. The *ANMRR* measure of the proposed scheme compared with other works for relevance feedback.

Relevance Feedback Algorithms	***ANMRR***
Method A	0.11
Method B	0.07
The Method of [9]	0.12
The Method of [8]	0.14
The Method of [7]	0.19