WEIGHT UPDATING FOR RELEVANCE FEEDBACK IN AUDIO RETRIEVAL

Mingchun Liu, Chunru Wan

School of Electrical and Electronic Engineering
Nanyang Technological University, Nanyang Avenue,Singapore 639798
{P147508078, Ecrwan}@ntu.edu.sg

ABSTRACT

The relevance feedback is proved to be an effective method in text information, image, and video retrievals. In this paper, we introduce this technique to carry out audio retrieval, in a hope not only to enhance the retrieval performance but also through this kind of user interaction to enhance the searching ability. Based on an initial searching result, a user can tag files with relevance or irrelevance according to one's judgment and preference. Then, the system updates the weights in similarity measurement and/or the query itself based on the feedbacks. Two relevance feedback algorithms have been proposed. One is a simplified technique used for feedback in image retrieval; another is based on constrained optimization concept. Experiments show that both approaches can yield similar performance improvements. Furthermore, the latter one can utilize negative feedbacks in a unified approach as well.

1. INTRODUCTION

The relevance feedback has been proved to be an effective method to increase performance in text retrieval [1], image retrieval [2] and video retrieval [3]. Among the literature, relevance feedback in image retrieval has been most intensively studied. In [4], a relevance feedback based interactive retrieval approach is proposed, which aims to narrow the gap between high-level concepts and low-level features and take advantage of the subjectivity of human perception of visual content. During the retrieval process, the user's high-level query and perception subjectivity are captured by dynamically updated weights based on the user's feedback.

Audio retrieval is a relative new branch of research in the context of content-based multimedia retrieval. However, when fully explored, it can also be very useful in many applications, such as audio database system and entertainment industry. A general audio classification and retrieval system is built by Wold, and et al [5]. In that system, sound is reduced to perceptual and acoustical features, in which users can search or retrieve sounds by different kinds of query. A new pattern classification method called the nearest feature line (NFL) is presented for the same kind of task and the resulting system is claimed to achieve better performance on a same audio database [6]. Among these audio retrieval systems, the user interaction is lacking. However, the user involvement may be crucial to achieve a better performance.

In this paper, we introduce two relevance feedback techniques to carry out retrieval in audio domain. The first method is similar to the feedback procedure proposed for image retrieval in [7] but it is simplified. The second method is proposed from a different perspective, which is based on constrained optimization.

The rest of paper is organized as follows. In Section 2, a general audio retrieval system is described briefly, including feature extraction, audio indexing and classification techniques. In Section 3, two relevance feedback algorithms are proposed. In Section 4, various experiments are carried out to test the performance of audio retrieval with relevance feedback. Finally, conclusions are given in Section 5.

2. CONTENT-BASED AUDIO RETRIEVAL

The process of a common audio retrieval system can be concisely divided into three parts: audio feature extraction, indexing or classification, and similarity measurement.

2.1 Feature Extractions and Normalization

The first step towards an audio retrieval system is feature extraction. The extracted feature vectors can represent audios and then these feature vectors are normalized for classification and indexing. Here, features are extracted from time, frequency and coefficient domains and they are combined to form the feature vector to represent the individual audio file.

Time domain features we use in the experiments include RMS (root mean square), ZCR (zero-crossing ratio), VDR (volume dynamic ratio) and silence ratio.

Frequency domain features include frequency centroid, bandwidth, four sub-band energy ratios, pitch, salience of pitch, spectrogram, first two formant frequencies, and formant amplitudes. First 13 orders of MFCCs (Mel-Frequency Cepstral Coefficients) are adopted as coefficient features.

Each audio feature is normalized over entire files in the database by subtracting its mean and dividing by its standard deviation. Normalization can ensure that contributions of all audio feature elements are adequately represented and prevent one feature from dominating the whole feature vector. Then, the audio is fully represented by its normalized feature vector. The details of the feature extraction can be found in [8].

2.2 Audio Indexing and Classification

Several deterministic and statistical classifiers such as nearest neighbor, modified k-Nearest Neighbor (k-NN), Gaussian Mixture Model (GMM) and neural network classifier have been used to classify the database. Thus, we can apply these classifiers to index the audio before the search begins. The benefit of this process is to reduce searching space by labeling the audio files or applying hierarchical search [9].

2.3 Audio Retrieval

When a user wants to retrieve some audio documents, he or she usually inputs a query example to the audio search engine and requests for finding relevant files to the query. A similarity measurement such as Euclidean distance between the query and sample audio files is computed. Then, a list of files based on the similarity distance is displayed to the user for listening and browsing.

3. RELEVANCE FEEDBACK

Based on the retrieval result, user can listen to the sounds and tag files with relevance or irrelevance according to his or her judgment and preference. Then, the system updates the result based on the feedback in order to find more relevant files in the user's point of view progressively. Basically, the purpose of relevance feedback is to move relevant files ranking to the top and irrelevant files ranking to the bottom. In principle, there are two ways to apply users' feedback strategy. One is to update the weights of similarity measurement and the other is to refine the query.

3.1 Relevance Feedback Algorithm I

This relevance feedback algorithm is similar to the algorithm proposed in [7]. The underlying concept of the algorithm can be often seen in other relevance feedbacks in image retrieval. The idea is that we should assign more weights to the feature, which has a diverse set of values in the whole database but similar value for relevant images. Then, more relevant images will appear on the top retrieval list in the next iteration of retrieval.

Based on this observation, we simplify the original algorithm in [7] and apply it in the audio retrieval. During the feedback process, the weighted *L2* distance instead of Euclidean distance calculates the similarity measurement. The weighted *L2* distance is defined as follows.

$$\rho(x, y : w) = \left(\sum_{i=1}^{N} w_i (x_i - y_i)^2 \right)^{1/2} \qquad (1)$$

where x and y are two given feature vectors, w is the weight vector and N is the number of features in the feature vectors.

The feedback retrieval algorithm is described as follows:

1. Initialize the weights w_i to $1/N$. That is,
$$w_i = 1/N, \quad i = 1,2,...N. \qquad (2)$$
2. Search the database using w_i and obtain retrieval result list using the Eq. (1).
3. Get feedback from the user and form the relevance audio set R_{rel}, the original query example is always included.
4. Calculate the standard deviation $\sigma_i = std(F_{rel,i})$, where $F_{rel,i}$ is the *ith* feature components of the audios in R_{rel}.
5. Update weights accordingly by the following rule:
$$w_i = (w_i + \Delta w_i) / (\sum_{i=1}^{N} \Delta w_i + 1) \qquad (3)$$
where $\Delta w_i = \frac{1}{\sigma_i + \alpha}$. Note that the standard deviation of the *ith* feature component of the whole audio database is *1* according to our normalization scheme. The constant α is an experimentally determined positive number. The sum of the new weights remains the same as the sum of old ones, which equals to *1*.
6. Go back to step 2 and search using the updated weights w_i.

3.2 Relevance Feedback Algorithm II

Suppose that we have obtained the relevance audios set R_{rel}, which includes the query example q and relevance feedbacks f^j, $j = 1, \cdots, M$. M is the number of relevant feedbacks. If we can decrease the sum of the square

weighted *L2* distance $\sum_{j\in R_{rel}} \rho^2(f^j, q, w)$ between relevance feedbacks and the query example, more relevant audios may emerge on the top of the retrieval list because of their similar feature characteristics. Based on this observation, we consider minimizing the following objective function:

$$f(w) = w^T D w + \varepsilon w^T w \quad \text{Subject to } c^T w = 1 \tag{4}$$

where ε is a positive constant,

$$D = diag\left\{ \sum_{j\in R_{rel}} d_{ji}^2 \right\} = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_N \end{bmatrix}, \mathrm{c} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{5}$$

Here d_{ji} is the distance between *ith* feature component of the *jth* relevance feedback and query example, defined as $d_{ji} = f_i^j - q_i$. Thus, $d_i = \sum_{j\in R_{rel}} d_{ji}^2$. (6)

The term $\varepsilon w^T w$ is introduced to avoid very large variation of w. This is a typical constrained optimization problem, which can be solved by the Lagrangian method. The solution of the constrained optimization problem is given by

$$w = \frac{R^{-1}c}{c^T R^{-1} c} = \frac{1}{(r_1^{-1} + \cdots + r_N^{-1})} \begin{bmatrix} r_1^{-1} \\ \vdots \\ r_N^{-1} \end{bmatrix} \tag{7}$$

or $w_i = \frac{1}{(r_1^{-1} + \cdots + r_N^{-1})} r_i^{-1}$ (8)

where $R = D + \varepsilon I$ and $r_i = d_i + \varepsilon$. In case of negative feedback, the objective function can be adjusted as follow:

$$\mathrm{f}(w) = w^T D w - \beta w^T D' w + \varepsilon w^T w + \lambda(c^T w - 1) \tag{9}$$

where β is a positive number and it is usually small to reduce the effect of negative feedback compared to positive feedback. $D' = diag\left\{ \sum_{j\in R_{irrel}} d'^2_{ji} \right\}$, R_{irrel} is defined as the irrelevance or negative feedback audio set. $d'_{ji} = f'^j_i - q_i$, f'^j is a negative feedback in the set R_{irrel}. In this case, the solution to Eq.(9) has the same form as in Eq.(7) and Eq.(8) with *R* being replaced by $R = D - \beta D' + \varepsilon I$.

4. EXPERIMENTS

In the retrieval system with feedback, different users may have different opinions and may choose different files as feedbacks or even determine the same file as relevance or irrelevance. In this paper, in order to avoid such evaluation problem, we conduct experiments in a fully automatic way. We assume files in the same classes in the database as relevant, otherwise as irrelevant.

4.1 Performance Evaluation

We use the same audio database as in [5], [6], [8]. There are 414 sound files all together, which collected from 16 classes, including female and male speeches, seven music instruments and seven environmental sounds. The precision and recall are frequently used as an effective measurement of retrieval performance. The precision is calculated by dividing the number of relevant retrieved files by the number of total retrieved files. The recall is computed by dividing the number of relevant retrieved files by the number of total relevant files. Sometimes, the average precision (AP) is used as one single measurement of retrieval performance, which refers to an average of precision at various points of recall. In most cases, however, users don't have patience to listen to all the possible retrieved files. Normally, they are only interested in several files ranking at the top. Thus, we calculate the top T (T=15) AP to indicate the practical retrieval performance. Since the file already have their labels, the performance can be measured automatically without hearing the sound. In our experiments below, we set $\alpha = 0.5, \beta = 0.1, \varepsilon = 0.5$.

4.2 Experimental Results

In the experiments, each audio file in the database is submitted to the search engine as query example one by one. The mean APs of the tests are measured and listed in Tables 1 and 2.

Firstly, the retrieval performance without feedback is measured. Since files in same class are already assumed as relevant, we can mark those files from most similar to least similar automatically. Therefore, the first 1-3 files are used as relevance feedback for weight updating and used in the next iteration of retrieval. In the meantime, we use the mean of files in the relevance audio set for query updating. The original mean AP is 0.485. From Table 1, we can see that if we chose first 3 relevant files as relevance feedback, the performance can increase to 0.577 and 0.59 by algorithm I and II respectively. In experiment, when the first top ranking irrelevance file is used as negative feedback in algorithm II, the AP can be further increased to 0.594.

Usually, people only browsing the files ranked on the top list. In Table 2, only the top 15 retrieved files are taken into consideration. The mean AP for the top 15 files is 0.807. It can increase to 0.923 and 0.93 by algorithm I and II in one feedback iteration with 3 relevance feedbacks. When the first top ranking irrelevance file is used as negative feedback in algorithm II, the AP can be further increased to 0.935.

In order to show the whole performance improvement rather than particular one, the AP differences after and

before feedbacks with 3 relevant files using algorithm II are shown in Figures 1 and 2. Figure 1 considers the whole retrieved files, while Figure 2 considers the top 15 retrieved files only. The bar above the horizontal line means that the AP after feedback is higher than before feedback and vice versa. We can clearly see that in most cases, the performances after feedbacks are better.

Table 1: Overall AP of the feedback experiment

Average-Precision	1st feedback with query updating	
	Algorithm I	Algorithm II
1 File	0.517	0.52
2 Files	0.551	0.558
3 Files	0.577	0.59

Table 2: Top T AP of the feedback experiment (T=15)

Average-Precision	1st feedback with query updating	
	Algorithm I	Algorithm II
1 File	0.855	0.857
2 Files	0.895	0.90
3 Files	0.923	0.93

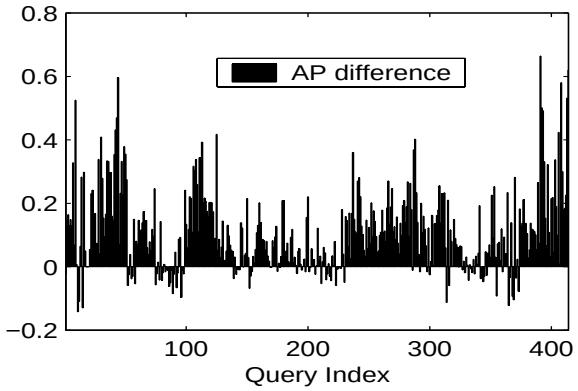

Figure 1: AP difference of each query performance

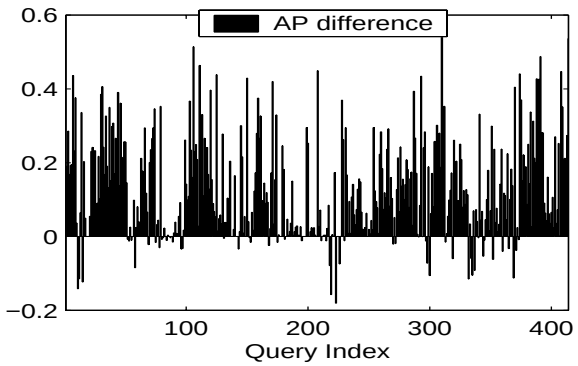

Figure 2: AP difference of each query performance with Top T files considered (T=15)

5. CONCLUSIONS

In this paper, two relevance feedback algorithms for weights updating in audio retrieval are proposed. The enhancement of retrieval ability by relevance feedbacks is demonstrated through experiments. Both algorithms have similar performance improvement with one iteration feedback. Yet, the second feedback algorithm, which is based on adaptive array processing and can also handle negative feedback, yields slightly better performance. This algorithm can also be applied in image retrieval using feedbacks. Through the relevance feedback, some intelligence or semantics can be added to the retrieval system thus the gap between the subjective concepts and objective features can be narrowed.

6. REFERENCES

[1] Chia-Hui Chang, and Ching-Chi Hsu, "Enabling concept-based relevance feedback for information retrieval on the WWW," *IEEE Transactions on Knowledge and Data Engineering*, pp: 595 –609, Vol: 11 Issue: 4, 1999

[2] Feng Jing, Mingiing Li, Hong-Jiang Zhang, and Bo Zhang, "Learning region weighting from relevance feedback in image retrieval," *IEEE ICASSP'02*, Vol: 4, pp: 4088 –4091, 2002

[3] Wang, R., Naphade, M.R., and Huang, T.S., "Video retrieval and relevance feedback in the context of a post-integration model," *IEEE Fourth Workshop on Multimedia Signal Processing*, pp: 33-38, 2001

[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, pp: 644 –655, 1998

[5] Wold E, Blum T, and Keislar D, et al, "Content-based classification, search, and retrieval of audio," *IEEE Multimedia*, pp. 27-36, Fall 1996

[6] S. Z. Li, "Content-based classification and retrieval of audio using the nearest feature line Method," *IEEE Transactions on Speech and Audio Processing*, Vol 8, Issue 5, pp: 619 –625, Sept 2000

[7] Aksoy, S., and Haralick, R.M., et al, "A weighted distance approach to relevance feedback," *Proceedings of 15th International Conference on Pattern Recognition*, pp: 812 -815 Vol.4, 2000

[8] Mingchun Liu, and Chunru Wan, "A study on content-based classification and retrieval of audio database," *Proceeding of International Database Engineering & Application Symposium*, pp.339-345, 2001

[9] Tong. Zhang, and C.-C. Jay Kuo, "Hierarchical classification of audio data for archiving and retrieving," *IEEE ICASSP'99*, Vol. 6, pp. 3001-3004, 1999