SYLLABLE-BASED RELEVANCE FEEDBACK TECHNIQUES FOR MANDARIN VOICE RECORD RETRIEVAL USING SPEECH QUERIES

Lin-Shan Lee[1,2], Bo-Ren Bai[1] and Lee-Feng Chien[2]

[1]Dept. of Electrical Engineering, National Taiwan University
[2]Institute of Information Science, Academia Sinica
Taipei, Taiwan, R.O.C.
E-mail: white@speech.ee.ntu.edu.tw

ABSTRACT

In order to solve the problem with the new environment of fast growth of audio resources on the Internet, we have presented a syllable-based approach which is capable of retrieving Mandarin voice records using queries of unconstrained speech[1]. However, the performance achieved by this previously proposed approach is still not satisfactory, and one of the reason is that very often the information provided by the speech query for the request subject may not be sufficient. In this paper, we present approaches based on the relevance feedback technique to improving the performances of the previous research. The proposed approaches include a relevance measure adjustment scheme using a relevance table for the voice database, a query expansion scheme to generate a new query including the feedback information, and a combination of these two schemes. Extensive preliminary experiments were performed and encouraging results were demonstrated.

1. INTRODUCTION

Due to the fast growth of audio resources published and broadcasted on the Internet, more and more information are stored in the form of voice databases. Techniques to retrieve desired information from a large voice database become very important. To retrieve voice message files using speech queries is very attractive, but this subject has not been adequately discussed in the literature, although some works of text retrieval of voice files[2][3] or voice retrieval of text files[4][5][6] have been reported. Some preliminary studies on retrieving Mandarin voice message files using queries of unconstrained speech have been recently reported[1]. By properly utilizing the monosyllabic structure of Chinese language, the previously proposed approach performed the complete matching process directly at the syllable level using syllable-based statistical information, and showed the feasibility to retrieve Mandarin voice files using speech queries.

However, the performance achieved by this previously proposed approach is still not satisfactory, and one of the reason is that very often the information provided by the speech query for the request subject may not be sufficient. Because the inclusion rates obtained in the previously proposed approach are usually high, it is thus believed that the precision rates of retrieval can be significantly improved by the concept of relevance feedback, which has been popularly used in the conventional text information retrieval (IR) area[7]. In other words, even if the results of the first retrieval are not good enough, it will be easy for the users to add extra information by feeding back some part of the previously retrieved results for further retrieval, and apparently better results can be attained after several times of feedback. Such a technique of relevance feedback has also been applied recently to voice record retrieval and shown to be very helpful[8]. In this paper, syllable-based relevance feedback approaches for Mandarin voice record retrieval using speech queries are developed and investigated.

Although there exist more then 10,000 commonly used Chinese characters, each character is monosyllabic and the total number of phonologically allowed Mandarin syllables is only 1345. The combination of these monosyllabic characters or 1345 syllables gives almost unlimited number of monosyllabic or polysyllabic Chinese words. This special monosyllabic feature of Chinese language makes it possible to compare the voice message records and the speech queries directly at syllable level. It was found that statistical parameters for syllable pair occurrence are very useful in such tasks and it is not necessary at all for retrieval purposes to go beyond the syllable to characters or words[1] such that the computation requirement can be significantly reduced. In this paper, the proposed relevance feedback approaches are also developed in the syllable level only. The proposed approaches include a relevance measure adjustment scheme using a relevance table for the voice database, a query expansion scheme to generate a new query including the feedback information, and the third approach combines the previous schemes to get better results. Extensive preliminary experiments were performed and encouraging results were demonstrated.

2. RELEVANCE TABLE

To implement the relevance feedback concept, a relevance table which records the relevance measure of the records in the database must be constructed first. A feature vector V_r for each voice record r in the database D is first defined:

$$V_r = (L_r(s_1,s_1) \times idf_D(s_1,s_1), ...,$$
$$L_r(s_i,s_j) \times idf_D(s_i,s_j), ..., \qquad r \in D \qquad (1)$$
$$L_r(s_{1345},s_{1345}) \times idf_D(s_{1345},s_{1345}))$$

where $L_r(s_i,s_j)$ is a likelihood measure which includes the occurrence frequency as well as the acoustic recognition scores for the syllable pair (s_i,s_j) obtained after pre-processing the voice record r, and $idf_D(s_i,s_j)$ denotes the inversion document frequency (IDF)[9] for the syllable pair (s_i,s_j) in the database D, which is defined as :

$$idf_D(s_i,s_j) \equiv log(N / N_{s_i,s_j}) \qquad (2)$$

where N is the total number of voice records in the database, and N_{s_i,s_j} is the number of records where the syllable pair (s_i,s_j) appears. The value of $idf_D(s_i,s_j)$ implies the significance of the syllable pair (s_i,s_j) in the database. The syllable pair with smaller $idf_D(s_i,s_j)$ value means to be less discriminative for information retrieval. For example, if a syllable pair (s_i,s_j) appears commonly in many voice records, N_{s_i,s_j} is large for the syllable pair, and we will give the syllable pair a small weight by the definition of $idf_D(s_i,s_j)$. This is very reasonable since a commonly-used vocabulary will be less important from the viewpoint of information retrieval.

The feature vectors V_r's have a dimension of 1345×1345, where 1345 is the total number of phonologically allowed tonal syllables in Chinese language, but practically the dimension can be reduced much smaller than this number by grouping technique[10]. Given the feature vectors V_r's for all the voice records r's in the database D, we can then define the relevance measure between every two records by the Cosine measure:

$$R(i,j) = cos(V_{r_i},V_{r_j}) = \frac{V_{r_i} \cdot V_{r_j}}{|V_{r_i}||V_{r_j}|} \qquad ,r_i,r_j \in D \qquad (3)$$

In this way, a larger value of $R(i,j)$ indicates the higher relevance between the record r_i and r_j. All these relevance measures are then used to construct a relevance table $R = \{R(i,j)|r_i,r_j \in D\}$.

3. INITIAL RETRIEVING PROCESS

In the retrieving process, the feature vector V_q for the input speech query q is first obtained in exactly the same way as the feature vector of a record V_r mentioned above. The relevance measure between each record r_j in the database D and the input query q, $R(q,j)$, is then defined in exactly the same way as $R(i,j)$ in equation (3) except V_{r_i} is replaced by V_q. An initial retrieval then simply select a set of records I_0 with higher relevance measure to the input query q.

A two-stage searching scheme has been proposed to initially retrieve some records which are more relevant the speech query[1]. The first stage is an approximate matching which is fast with high recall rate. In order to delete most irrelevant records, here we just compare the acoustic score information of the syllable pairs in the feature vectors for the query with those for the voice records. The second stage then performs detailed matching based on a specially designed similarity measure between the input query and the selected records by using the accumulated acoustic score and IDF information to provide a better result. It has been proved that the two-stage searching scheme is capable of providing high inclusion rates for retrieving by both simple queries and quasi-natural-language queries. Such that it is possible to further improve the results by applying the relevance feedback techniques.

4. RELEVANCE FEEDBACK

Now, given the results of the initial retrieval, to obtain better retrieving results, the user can further pick up a set of most related records I out of I_0 and feed I back to the system for further retrieval. With the extra information I provided by the user, the system can apparently retrieve more relevant records. This is the basic concept of relevance feedback.

Three approaches were developed to use the feedback information I for further retrieval. The first approach adopts a relevance measure adjustment scheme, the second one adopts a query expansion scheme, and the third one is a combination of the previous two schemes. The block diagram of the proposed approaches is shown in Figure 1, where the shadowed part can be off-line processed in advance to get the feature vector database and the relevance table, and the wide arrows represent the data flows of the feedback information.

4.1 Relevance Measure Adjustment

In this approach, we try to adjust the scores of the relevance measure of the records to the query under the help of the extra information from the feedback records. In Figure 1, the data flow of the feedback information for this approach is represented by

wide arrows of light gray. To perform this approach, a set of new records K highly related to the records in I are first identified, i.e., each record $r_k \in K$ has a relevance measure $R(i,k)$ exceeding a threshold for at least one record $r_i \in I$. Only this new set of records are then included in the further retrieval. The relevance measure between each record $r_k \in K$ and the input query q, $R(q,k)$, can then be adjusted into $R^*(q,k)$ by adding all values of $R(q,i)$ weighted by $R(k,i)$:

$$R^*(q,k) = R(q,k) + \sum_{i \in I} R(q,i) * R(k,i) \qquad k \in K \qquad (4)$$

The records r_k with higher relevance measure $R^*(q,k)$ can then be picked up as the new results.

4.2 Query Expansion

The second approach adopts a query expansion scheme. In this approach, some significant information in the results of the initial retrieval are extracted and added to the feature vector of the original input query to form a new query feature vector. Since this new query feature vector contains much more relevant information than that in the initial query feature vector, better results can thus be achievable. In Figure 1, the date flow of the feedback information for this approach is represented by the wide arrows of dark gray. The new feature vector V_q^* adopted in this work is defined as :

$$V_q^* = (u_1, u_2, \cdots, u_m, \cdots u_M) \qquad (5)$$

where u_m is the $m-th$ element of V_q^* and $M = 1345 \times 1345$ is the dimension of the feature vector, and

$$u_m = max(u_{qm}, u_{1m}, u_{2m}, \cdots, u_{im}, \cdots u_{Nm}) \qquad (6)$$

where u_{qm} is the $m-th$ element of V_q, and u_{im} is the $m-th$ element of V_{r_i}, $r_i \in I$, $i = 1,2,\cdots,N$, N is the total number of records in I selected by the user. With this new query feature vector feedback to the system, better results can be obtained in the retrieving process, since all important feature parameters are further emphasized.

4.3 Combined Approach

The third approach combines both relevance measure adjustment and query expansion schemes. That is, for records K highly related to the feedback records in I, the equation (4) is first performed to give each of these K records a new score, and the new query feature vector described in Section 4.2 is also used to give each of these K records a new score by performing

equation (3), these two kinds new scores then are combined to give each record a new relevance measure. All wide arrows in Figure 1 represent the date flows of feedback information used in this approach.

5. EXPERIMENTAL RESULTS

Extensive preliminary experiments have been performed in this work. A database including about 1,000 Mandarin voice records is first collected and processed to obtain feature vectors for the records, and the relevance table is also constructed in advance. For each record, a subject sentence is also assigned and included to describe its subject. Given an input query, a set of records I_0 relevant to the query was selected first and their subject sentences were played. Based on the subject sentences, the user then pick up several most relevant records I from I_0. The relevance feedback approaches were then performed to obtain more relevant records. Figure 2 shows the performance of the relevance measure adjustment scheme, the curves in the figure are the precision rates with respect to the inclusion rates from zero to two times of relevance feedback. Similar performance curves of the query expansion scheme and the combined approach are shown in Figure 3 and Figure 4 respectively. These results show that the precision rates will get better and better with more and more times of feedback. And Figure 4 show that better results can be achieved by the combination of different schemes.

6. CONCLUSION

In this paper, we present approaches based on the relevance feedback techniques to improving the performance of retrieving Mandarin voice records using speech queries. These approaches can add extra information .by feeding back some part of the previously retrieved results for further retrieval and improve the precision rates of the retrieving results with several times of feedback.

References

[1] Bo-Ren Bai, Lee-Feng Chien, and Lin-Shan Lee, "Very-Large-Vocabulary Mandarin Voice Message File Retrieval Using Speech Queries", *Proc. Inc. Conf. on Spoken Language Processing*, Vol. 3, pp. 1950-1953, 1996.

[2] U. Glavitsch and P. Schauble, "A System for Retrieval Speech Documents", *ACM SIGIR Conference on R&D in Information Retrieval*, pp. 168-176, 1992.

[3] K. Sparck Jones, G. J. F. Jones, J. T. Foote and S. J. Young, "Experiments in Spoken Document Retrieval", *Information Processing and Management*, Vol. 32, No. 4, pp. 399-417, 1996.

[4] Sung-Chien Lin, Lee-Feng Chien, Keh-Jiann Chen, and Lin-Shan Lee, "Unconstrained Speech Retrieval for Chinese Document Databases with Very Large Vocabulary and Unlimited Domains", *Proc. European conf. On Speech Commun. And Techno.*, Vol. 2, pp. 1203-1206, 1995.

[5] Sung-Chien Lin, Lee-Feng Chien, Keh-Jiann Chen, and Lin-Shan Lee, "An Efficient Voice Retrieval System for Very-Large-Vocabulary Chinese Textual Databases with a Clustered Language Model", *Proc. 1996 IEEE Int. Conf. on Acoust., Speech and Signal Processing*, Vol. 1, pp. 287-290, 1996.

[6] Julian Kupiec, Don Kimber and Vijay Balasubramanian, "Speech-Based Retrieval Using Semantic Co-Occurrence Filtering", *Proc. of The Human Knowledge Technology Workshop*, pp. 373-377, 1994.

[7] Donna Harman, "Relevance Feedback and Other Query Modification Technique", Information Retrieval, Chap. 11, by Prentice Hall, 1992.

[8] David A. James, "A System for Unrestricted Topic Retrieval from Radio News Broadcasts", *Proc. 1996 IEEE Int. Conf. on Acoust., Speech and Signal Processing*, Vol. 1, pp. 279-282, 1996.

[9] G. Salton, Introduction to Modern Information Retrieval, NY, McGraw-Hill, 1983.

[10]Lee-Feng Chien, "Fast and Quasi-Natural Language Search for Gigabytes of Chinese Texts", *ACM SIGIR Conference on R&D in Information Retrieval*, pp. 112-120, 1995.
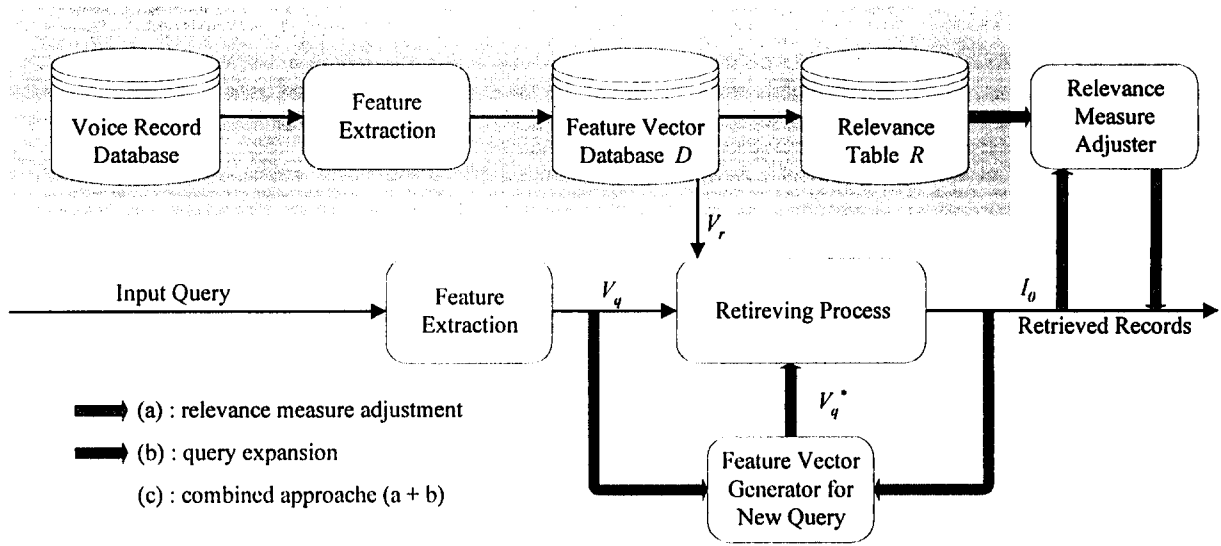
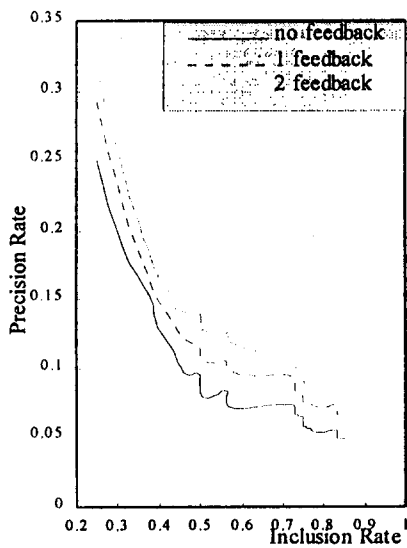Figure 1. The block diagram of the three relevant feedback approaches.
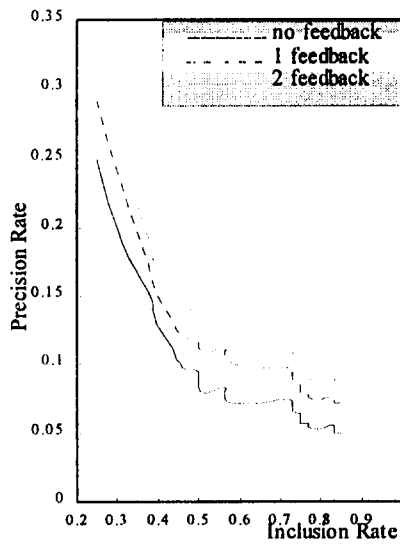

Figure2. Performance of relevance measure adjustment
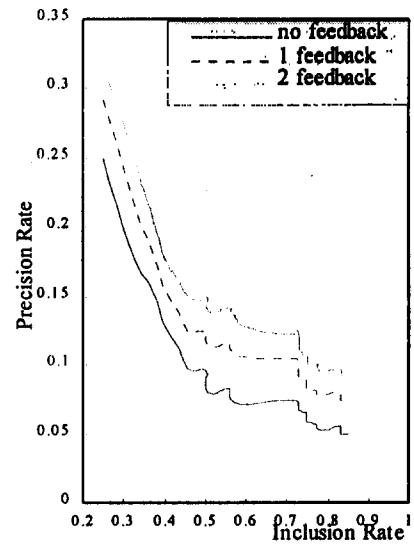
Figure3. Performance of query expansion

Figure4. Performance of the combined approach