Distributed Fuzzy Rules for Preprocessing of Speech Segmentation with Genetic Algorithm

Ching-Tang Hsieh, Eugene Lai and You-Chuang Wang
Department of Electrical Engineering,
Tamkang University, Taipei , Taiwan,
Republic of China

Abstract

Most of the speech segmentation works are based on the thresholds of parameters to segment the speech data into phonemic units or syllabic units. In this paper, we formulate the threshold decision as a clustering problem. Feature parameters extracted from the analysis frame are clustered into three types: silence, consonants, and vowels. Distributed fuzzy rules which have been used in clustering the numerical data are used for this task. The distributed fuzzy rules, which do not need many training data, have good performance in clustering problem and are beneficial for clustering the features of speech data. Such method, however, has many fuzzy if-then rules. So, we propose a genetic-algorithm-based method for selecting a small number of significant fuzzy if-then rules to construct a compact fuzzy classification system with high classification power. Effectiveness of this approach has been substantiated by classification experiments for continuous radio news speech samples uttered by two females and two males.

keyword : speech segmentation; distributed fuzzy rules; genetic-algorithm.

1. Introduction

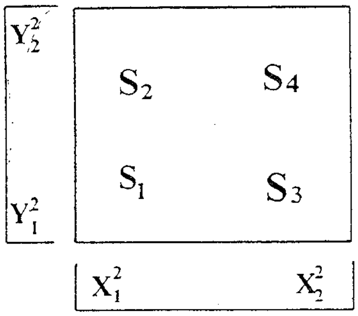
The requirement of memory space and the complexity of mathematical operations in large - vocabulary continuous speech recognition system, speech segmentation is an important preprocessing work. In general, there are two kinds of segmentation. One is phonemic unit segmentation [1,2], which segments the speech into phonemes based on the features of phonemes. The other is syllabic unit segmentation[3,4]. Most of the syllabic unit segmentation which locates the endpoints of an utterance are based on two measures of the signal, zero crossing rate and energy. For either phonemic unit segmentation or syllabic unit segmentation, the choice of thresholds of segmentation parameters is a basic and an important work. However, most of the thresholds, which have been simply used for deciding voiced / unvoiced (the average value of first formant band [1,2]) or consonants/ silence (zero crossing rate and energy [3,4]) , are found by personal experience such that there exists some trade-offs in the decision. e.g. Choosing higher threshold of energy causes losing error. Setting lower threshold of energy causes adding error. By this consideration, a preprocessing of speech segmentation is necessary to classify speech into silence, consonant and vowel. We formulate the decision of thresholds as clustering problem to avoid inadequate settings. The clustering problem is solved by the fuzzy methodology. One of them, which was proposed by Hisao, Ken, and Hideo[6], is the "Distributed Fuzzy Rules". Without many training data, the correct ratio for clustering unknown samples by the distributed fuzzyrules is high. This property is beneficial to scope the features of utterances. However, when we deal with speech classification problems, we always need a large number of fuzzy if-then rules to achieve the high classification ability. Such process wastes time and can't be executed on a real-time system. Recently, genetic algorithms are used extensively in many fields and its optimal results are also satisfied by us. Hideo used triangular membership function for training fuzzy rules and selecting fuzzy rules by GA[7]. Owing to the character of speech signl, we use exponential membership function. Therefore, this paper proposes a genetic-algorithm-based approach to achieve the high classification ability and construct a minimal set of the fuzzy if-then rules. Our experimental results seem encouraging this approach. The remainder of the paper is organized as follows. Section II briefly describes the distributed fuzzy rules and genetic algorithm. In Section III, we introduce the parameters for phonemic features and classes. Section IV is our simulation result. Finally, concluding remarks are presented in Section V.

2. Phonemic Features Clustering Using Fuzzy Rules and Genetic Algorithm

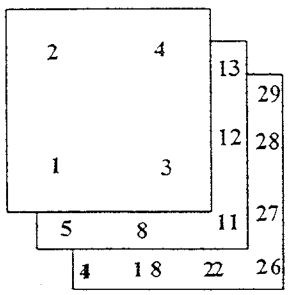
2.1 Distributed Fuzzy Rules

Hisao, Ken, and Hideo[6] proposed the "Distributed Fuzzy Rules" to cluster the numerical data using the triangular membership function. For only 3 classes and 9 training samples, the correct ratio for clustering unknown samples is up to 90%. There are two conclusions: (1) Under the same fuzzy partition, the correct ratio by

0-7803-3796-4/97/$10.00©1997IEEE



(a) Labels and indices of fuzzy if-then rules (L=2)



(b) Distributed Fuzzy Rules (L=4)

Fig 1. The representations of different fuzzy rules

distributed fuzzy rules is higher than by the ordinary fuzzy rules; (2) Even since fewer training samples, the correct ratio by distributed fuzzy rules to cluster unknown samples is still higher. These properties are beneficial for clustering the features of large speech data without many training data. Because of the character of speech signals, we use the following exponential membership function

$$\mu_i^k(x) = \exp(-\beta^2(x - a_i^k)^2) \quad i = 1, 2, \ldots\ldots, k \ (k \geq 2) \quad (1)$$

where $\mu_i^k(x)$ is the membership function of subspace A_i^k and

$$a_i^k = (i - 1) / (k - 1), \qquad i = 1, 2, \ldots, k, \qquad (2)$$

The ordinary fuzzy rules can be described as

if x is X_i^l and y is Y_j^l

then [x,y] belongs to $[X_i^l, Y_j^l]$;

$i, j = 1,2, \ldots\ldots, L$ (3)

where [x , y] is one sample in unit square [0,1]x[0,1] ,and $[X_i^l, Y_j^l]$ is the subspace of unit square, and L is the fuzzy partitions. Making some modifications of eq. (3) as

if x is X_i^k and y is Y_j^k

then [x,y] belongs to $[X_i^k, Y_j^k]$;

$i, j = 1,2,\ldots,k;\ k = 2,3,\ldots,L$ (4)

becomes the distributed fuzzy rules. Different number of fuzzy partitions is the difference between ordinary fuzzy rules and distributed fuzzy rules. For L partitions, the distributed fuzzy rules are operated on L-1 unit spaces, i. e. from 2 partitions to L partitions. The ordinary fuzzy rules are operated only one unit space, i. e. the unit space is with L partitions. Fig. 1 shows the representations of different fuzzy rules. Clustering every subspaces from the training data to obtain eq. (5) is the training purpose.

if x is X_i^k and y is Y_j^k

then [x,y] belongs to class-x (5)

In eq(5), class-x is the class with own data; the class with maximum membership among all subspaces classes is the class of unknown data. More details can be seen from reference[6].

2.2 Genetic Algorithm

We can get the set of fuzzy rules by training some speech data. Our problem in this section is to select fuzzy rules from all the fuzzy rules to construct a compact rule set with high classification power. We can briefly describes the Genetic algorithm operations as follows[7]:

i) Initialization: Generate some initial populations randomly that contain some string ($S=S_1S_2\ldots\ldots S_N$) where

N is the total number of fuzzy rules.

S_n=1 denotes that the nth rule is selected.

S_n=-1 denotes that the nth rule is not selected.

S_n=0 denotes that the nth rule is a dummy rule.

ii) Fitness: Our purpose is to maximize the number of correctly classifield speech data by selected fuzzy rules set S and to minimize the number of fuzzy rules in S. So, the fitness value of each string can be formulated as

$$F(s) = W_{NOC} \bullet NOC(s) - W_S \bullet |S| \qquad (6)$$

where

$F(s)$ is the fitness value($F(s) \geq 0$), NOC(s) is the number of correctly classified speech data by S and |S| is the number of fuzzy rules in S. W_{NOC} and W_s are positive weights.

iii) Reproduction: The selection probability of the individual S in new generation is proportional to its fitness value

iv) Crossover: We apply one point for crossover to the pair of selected individuals such that we can get new strings.(This step repeats p/2 times, where p is the number of populations).

v) Mutation: This operation can prevent strings to locate into local optimization.

$$S_n \rightarrow S_n \times (-1) \tag{7}$$

Repeat steps iii, iv and v until satisfy the stopping condition.

3. The Parameters for Phonemic Features and Classes

Two parameters, the average value of first formant band (Vi) and zero crossing rate. are used for clustering. The first parameter is based on the spectral envelopes of speech obtained by the unbiased log spectral estimation [5]. This parameter has been used in voiced/unvoiced decision, and has achieved good performance [1,2]. It represents the degree of period of vowels. The second parameter, usually be used to decide the startpoints and endpoints of syllables [3,4], represents the degree of fricatives. These two parameters are in charge of the two axes of fuzzy space to construct the distributed fuzzy rules.

The structures of Mandarin speech are different from the other languages, such as English. Basically, the structures of Mandarin words can be expressed as

[Consonant,][Median,]Vowel[,Tail], (8)

where [] is optional. In Mandarin, the Median may be pronounced as /i/, /u/, or /iu/.The Tail may be pronounced as /n/ or /ng/. Since Median and Tail are at different positions with vowels, they are the phonemes extracted from vowels fron acoustic view point. Thus, we can reconstruct the structures of Mandarin words as

[Consonant,]Vowel. (9)

Since the structures of Mandarin words may be expressed by eq. (9), we can conclude that the speech of Mandarin can be classified into three classes: (1) silence-class; (2)consonant-class; and (3)vowel-class. Table I shows the relationship between phonemes and the classes.

Table I. Relationship between Classes and Phonemes

Class	Phoneme
Silence	silence
Consonants	f,d,t,g,k,h,j,b,p,y,m,ch,sh,tz,ts,s,r,l
Vowels	i,u,a,o,e,io,ai,ei,au,ou,el,ia,ie,iau,iou,ua, uo,uai,uei,ue,iue,iua,iu,n,ng

4. Experiments for classification

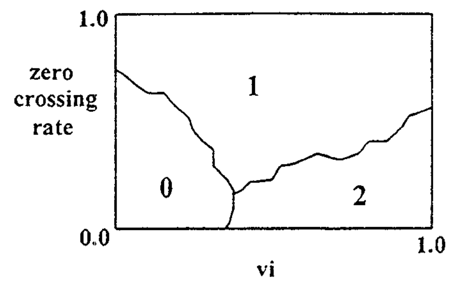


Fig 2. The distribution of classes for 20 training blocks

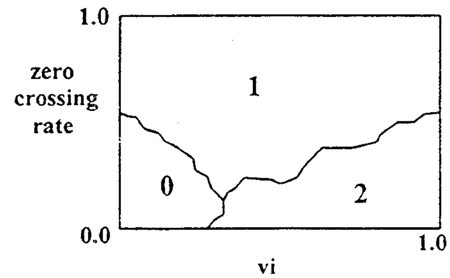


Fig 3. The distribution of classes with GA for 20 training blocks
(Using triangular membership function)

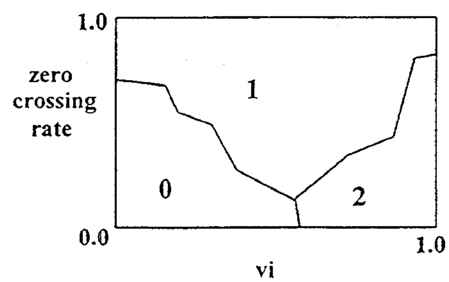


Fig 4. The distribution of classes with GA for 20 training blocks
(Using exponential membership function)

Table II. The parameters of Genetic Algorithm

Population size (N_{POP})	20
Number of generations	50
Crossover probability (P_C)	0.9
Mutation probability (P_m)	0.02

Experimental speech data of 2 female and 2 male speakers are sampled from the news on the radio. Fig.2 shows the regions of silence , consonants, and vowels classes which came from 20 different numbers of block of training data (4 seconds in each block). The genetic operations in section 2.2 are applied to this problem. We use the triangular membership and exponential membership to train the fuzzy if-then rules, the selected fuzzy if-then rules after 50 generations and the corresponding classification results are shown in Fig 3 and Fig 4. In computer simulation, the parameters of genetic algorithm are shown in Table II. The results of classification by distributed fuzzy rules using triangular membership function and exponential membership function are shown in Table III and Table IV, respectively. From Table III, we can see that the classification rate of Consonant is very bad when L=6 and even we use more fuzzy rules by increasing L, the classification result is improved limitedly. So, it must exist many dummy rules. From Table IV, we can see that the classification rate of Silence, Consonant and Vowel are high and changed with different value of β ,even though it has only 90 fuzzy if-then rules.

Table III. Classification results with different values of L by distributed fuzzy rules using triangular membership function.

	L=6	L=12	L=13	L=14	L=15
Silence	94.35%	84.85%	84.85%	84.77%	85.21%
Consonants	70.11%	83.54%	83.77%	83.84%	84.14%
Vowels	98.76%	96.02%	95.95%	95.91%	95.87%
average	87.74%	88.14%	88.19%	88.17%	88.41%
fuzzy rules	90	649	818	1014	1239

Table IV. Classification results with different values of β by distributed fuzzy rules using exponential membership function(L = 6).

	$\beta = 6$	$\beta = 7$	$\beta = 8$	$\beta = 9$	$\beta = 10$
Silence	88.70%	88.24%	87.77%	87.38%	87.23%
Consonants	79.75%	82.04%	82.96%	83.40%	83.70%
Vowels	98.05%	95.66%	95.30%	94.90%	94.68%
average	88.83%	88.65%	88.68%	88.56%	88.54%

Table V and Table VI show the classification result by genetic algorithm. Compared with Table III and Table V, we know that even fuzzy rules decrease to 1/3, the classification result almost doesn't change. Compare with Table IV and Table VI, we can see even fuzzy rules decrease to 1/5, the classification result almost become worse. From these observations, we have the following conclusions: (1) Without many training data, the clustering results are almost the same.This is the property of distributed fuzzy rules [6]; (2)A genetic-algorithm-based method for selecting significant fuzzy if-then rules can construct a compact fuzzy classification system (3)Using exponential

Table V. Classification results with different value of L by genetic algorithm.
(Using triangular membership function)

	L=6	L=12	L=13	L=14	L=15
Silence	89.93%	81.13%	83.44%	87.87%	83.52%
Consonants	78.63%	85.17%	85.84%	81.25%	85.47%
Vowels	96.63%	96.79%	95.20%	96.89%	95.14%
average	88.40%	87.70%	88.16%	88.67%	88.04%
fuzzy rules	38	239	314	384	485

Table VI. Classification results with β =10, L=6 by genetic algorithm.
(Using exponential membership function

	Silence	Consonant	Vowel	fuzzy rules
1	83.13	85.62	94.19	20
2	90	80.43	96.01	21
3	87.15	82.28	95.48	15
4	90.56	80.06	95.35	17
5	90.25	80.37	96.06	18
average	88.22	81.75	95.42	18.2

membership function is better than triangular membership function. (4)Vowels belonging to the vowel-class are up to 95%. The vowel-class can represent the feature of Vowels. (5) Consonants in the consonant-class is about 81%. Most of the consonants, especially the fricatives, can be considered to be identified by the consonant-class. (6)Silence belonging to the silence-class is 88% or so. We use the silence-class to represent the silence data.

5. Concluding Remarks

In this paper a preprocessing of speech segmentation is done and phonemic features are formulated as a clustering problem. The distributed fuzzy rules with genetic algorithms have been used successfully and the classification result doesn't become worse when we decrease the fuzzy rules. There are two advantages for this approach. First, the thresholds for classification work are found by clustering procedure to avoid inadequate setting. Second, during the clustering procedure, the labels which represent the features of phonemes are given for the speech data. This information can provide not only for the classification work but also can be used in the the consonants, vowel, and silence messages for the further recognition work. After the preprocessing of speech segmentation, we need to construct mandarin word from the labeling sequence by utilizing the structures mandarin words.

Acknowledgments:

This work was supported by the National Science Council (NSC), Taiwan, R. O. C., under the Grant NSC 85-2213-E-032-010

Reference

[1] Ching-Tang Hsieh and Jen-Tzung Chien," Seg menta tion of Continuous Speech into Phonemic Units,"ICIS. 1991, pp. 420-424.

[2] Heng-Kuan Liao, "Phonemic Units Segmentation and Rec ognition for Mandarin Chinese in Various Phonetic Environments,"M.S. Thesis,Tamkang University, June, 1993.

[3] L.R. Rabiner and M.R. Sambur, "An Algorithm for De terming the Endpoints of Isolated Utterances," The Bell System Technical Journal. vol. 54, no. 2, Febru ary 1975, pp. 297-315.

[4] Rong-Guey Chen,"Automatic Segmentation Techiques for Mandarin Speech Recognition," M.S. Thesis, National Taiwan University, June, 1978.

[5] Satoshi Imai and Chieko Furuichi, "Unbiased Esti mator of Log Spectrum and Its Application to Speech Signal Processing," Trans. IEICS, vol. J 70 -A, no.3, March 1987, pp. 471-480.

[6] Hisao Ishibuchi, Ken Nozaki and Hideo Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," Fuzzy Sets and System vol. 52,1992,pp. 21-31.

[7] Hisao Ishibuchi, Ken Nozaki and Hideo Tanaka, "Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms,"IEEE Trans. Fuzzy Systems, vol. 3, no.3, pp. 260-270, 1995.

[8] Ching-Tang Hsich and Shih-Chich Chien,"Speech Seg men tation and Clustering Problem Based on Fuzzy Rules and Transition states," IASTED International Conference on Applied Informations, annecy, May 1994,Annecy,Frency pp.291-294 .