THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.20, No.2E 2001. 6. pp. 17~24

Text-independent Speaker Identification by Bagging VQ Classifier

YounJeong Kyung*, BongDae Choi**, HwangSoo Lee***

*Platform R&D Center, SK Telecom
**Department of Mathematics, Korea University
***Department of EECS, Korea Advanced Institute of Science and Technology

(Received 27 October 2000 ; accepted 21 May 2001)

Abstract

In this paper, we propose the bootstrap and aggregating (bagging) vector quantization (VQ) classifier to improve the performance of the text-independent speaker recognition system. This method generates multiple training data sets by resampling the original training data set, constructs the corresponding VQ classifiers, and then integrates the multiple VQ classifiers into a single classifier by voting.

The bagging method has been proven to greatly improve the performance of unstable classifiers. Through two different experiments, this paper shows that the VQ classifier is unstable. In one of these experiments, the bias and variance of a VQ classifier are computed with a waveform database. The variance of the VQ classifier is compared with that of the classification and regression tree (CART) classifier[1]. The variance of the VQ classifier is shown to be as large as that of the CART classifier. The other experiment involves speaker recognition. The speaker recognition rates vary significantly by the minor changes in the training data set.

The speaker recognition experiments involving a closed set, text-independent and speaker identification are performed with the TIMIT database to compare the performance of the bagging VQ classifier with that of the conventional VQ classifier. The bagging VQ classifier yields improved performance over the conventional VQ classifier. It also outperforms the conventional VQ classifier in small training data set problems.

Keywords: *Speaker recognition, Text-independent, Vector quantization, Bootstrap, Variance*

I. Introduction

Many different models as like dynamic time warping (DTW), vector quantization (VQ), hidden Markov model (HMM) and Gaussian mixture model (GMM) and so on, exist for speaker recognition. The DTW can not applied to text-independent system though has a good performance in text-dependent system. The VQ is easy to construct and has a good performance. Also it has a common structure for speaker identification (SI) and speaker verification (SV). Temporal variation in speech signal parameters over the long term can be represented by stochastic Markovian transitions between states. The HMM has a good recognition

Corresponding author: YounJeong Kyung (yjkyung@sktelecom.com)
Platform R&D Center, SK Telecom,
99 Surin-Dong, Chongno-Gu, Seoul, 110-110, Korea

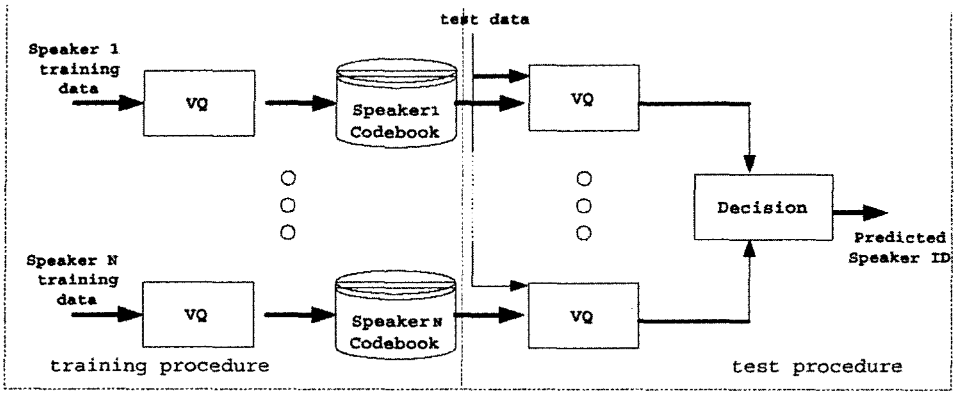

Figure 1. The conventional VQ classifier.

rates. However, when little data is available, the VQ-based method is more robust than a continuous HMM method [2]. The speaker is represented some Gaussian mixtures in GMM. The individual component Gaussians in a speaker dependent GMM are shown to represent spectral structure that is characteristic of easily recognized broad phonetic classes. But this system needs large training data. At least 30 seconds' speech is required for training[3]. As a result of model comparisons, we choose the VQ model for our speaker recognition system.

The VQ is a source coding technique that has been used successfully in both speech coding and speech/speaker recognition. In the VQ, each source vector is coded as one of a prestored set of codewords, called a codebook. The codebook is designed to minimize the average quantization distortion between itself and the training sequence.

To use VQ source coding in speaker recognition, we represent each speaker by a VQ codebook designed from a training sequence composed of repetitions of a particular utterance. We use the average quantization distortion in making the speaker identification decision. That is, VQ codebooks consisting of a small number of representative feature vectors are used as an efficient means of characterizing speaker-specific features. Clustering the training feature vectors of each speaker generates a speaker-specific codebook. In the recognition stage, an input utterance is vector-quantized using the codebook of each reference speaker, and the VQ distortion accumulated over the entire input utterance is used to make the recognition decision[4-6].

The diagram of this conventional VQ classifier for ASR (Automatic Speaker Recognition) is shown in Figure 1.

In this paper, we apply the bagging method to the VQ classifier in order to improve the recognition rates of ASR system. A bagging method, an acronym for "bootstrap and aggregating", is one of the perturb and combine (P & C) methods.

The bagging method generates multiple versions of a classifier by bootstrapping the training set and then combines these multiple versions into a single classifier by voting. We implement the ASR system by the bagging VQ classifier.

The construction algorithm of the bagging VQ is presented, and then the performance of the conventional VQ classifier is compared with that of the bagging VQ classifier. The bagging VQ classifier with a small training database is shown to have the better performance.

In [1,7], the bagging method works well for unstable classifiers such as the classification and regression tree (CART) and neural net. That is, it is known that the bagging method notably improves the performance of unstable classifiers. Two experiments will show that the VQ classifier is unstable. One of them is computation of the VQ classifier's variance with a waveform database[8].

We compare the variance of VQ classifier with that of CART[1]. The variance of the VQ classifier is as large as the CART's variance. The other experiment is the ASR test using a VQ classifier with minor changes in the training data set. Our experimental results confirm the unstability of VQ classifier. Thus, the bagging method is expected to improve the VQ classifier performance.

The paper is organized as follows: Chapter 2 presents a brief review of the bagging method and describes the notions of the bias and variance. In Chapter 3, we consider the unstability of the VQ classifier. The bias and variance of the VQ classifier are computed in this Chapter. These values are compared with those of CART. Chapter 4 presents the algorithm of the bagging VQ classifier for ASR. We give the experimental results in Chapter 5. Finally, Chapter 6 summarizes our results.

II. Bootstrap and aggregating

2.1. Review of the bagging method

The concept of the bagging method is as follows: A classification method is unstable if small perturbations in their training sets or in their construction can result in large changes in the constructed classifier. Unstable classifiers can have their accuracy improved by perturb and combine methods. That is, multiple versions of the classifier are generated by perturbing the training set of the construction method, then these multiple versions are combined into a single classifier[7,9].

2.2. Bias and variance of classifier

Breiman introduced the notion of classifier bias and variance[1].

Let the training data set L consist of data $(\boldsymbol{x}, y)$ where y is the class label of classifier $C(\boldsymbol{x}, L)$ if the input is $\boldsymbol{x}$, we predict y by $C(\boldsymbol{x}, L)$. Usually we have a single training data set L. Take repeated bootstrap sets $\{L_b\}$ from L, and form classifiers $\{C(\boldsymbol{x}, L_b)\}$.

Let

$$Q(y \mid \boldsymbol{x}) = P(C(x, L) = y) \tag{1}$$

and define the aggregated classifier as

$$C_A(\boldsymbol{x}) = \arg\max{}_y Q(y \mid \boldsymbol{x}). \tag{2}$$

This is aggregation by voting. In other words, consider many independent training data sets L_1, L_2, ...; construct the classifiers $C(\boldsymbol{x}, L_1)$, $C(\boldsymbol{x}, L_2)$...; and at each $\boldsymbol{x}$ determine the classification $C_A(\boldsymbol{x})$ by voting these multiple classifiers to find the most popular class.

The Bayes theorem is

$$P(y \mid \boldsymbol{x}) = \frac{P(\boldsymbol{x} \mid y)P(y)}{P(\boldsymbol{x})}, \tag{3}$$

where $P(y \mid \boldsymbol{x})$ is the probability of y, given $\boldsymbol{x}$. According to Equation (3), the Bayes classifier C^* is defined as

$$C^*(\boldsymbol{x}) = \arg\max{}_y P(y \mid \boldsymbol{x}). \tag{4}$$

That is, the Bayes classifier C^* chooses the class label y, which maximize the probability $P(y \mid \boldsymbol{x})$. The classifier $C(\boldsymbol{x})$ is unbiased at $\boldsymbol{x}$, if the predicted class label y of the aggregated classifier is not the same as the Bayes classifier. That is, we get the definition,

$$C(\boldsymbol{x}) \text{ is unbiased at } \boldsymbol{x} \text{ if } C_A(\boldsymbol{x}) \neq C^*(\boldsymbol{x}). \tag{5}$$

Let U be the set of all $\boldsymbol{x}$ at which C is unbiased, and call U the unbiased set. The complement of U is called the bias set and denoted by B. Each element of the other set T is assigned to either the bias set or unbias set.

Breiman introduced the bias and variance of classifiers as follows: The bias of a classifier C is

$$\begin{aligned} Bias(C) = P_{x,y}(C^*(\boldsymbol{x}) = Y, \boldsymbol{x} \in B) \\ - E_T P_{x,y}(C(\boldsymbol{x}, T) = Y, \boldsymbol{x} \in B) \end{aligned} \tag{6}$$

and its variance is

$$\begin{aligned} Var(C) = P_{x,y}(C^*(\boldsymbol{x}) = Y, \boldsymbol{x} \in U) \\ - E_T P_{x,y}(C(\boldsymbol{x}, T) = Y, \boldsymbol{x} \in U). \end{aligned} \tag{7}$$

If the variance is large, the classifier is highly affected by construction of training data sets. On the contrary, if the variance is small, the classifier is stable on construction of training data sets. That is, the variance of the classifier measures the unstability of the classifier.

III. Unstability of the VQ classifier

The main effect of the bagging method is to reduce the variance. Therefore a classifier's unstability is a key determiner for applying the bagging method. Unstable classifiers are characterized by high variance[7]. As training set L changes, the classifiers $C(\boldsymbol{x}, L)$ can differ significantly from each other and from the aggregated classifier $C_A(\boldsymbol{x})$. Stable classifiers do not change much over replicates of L, so $C(\boldsymbol{x}, L)$ and $C_A(\boldsymbol{x})$ will tend to be the same and the variance will be small.

Two experiments show that the VQ classifier is unstable. One is to compute the variance of the VQ classifier with a waveform database and show that its variance is large. The other one is to perform the ASR using the VQ classifier with minor changes to the training data set and show that the recognition results significantly vary depending on the training data set.

We compute the bias and variance with a waveform database[8]. Breiman computed variances of CART and, bagging CART in[1] with the waveform database. We use the same simulated database "waveform" to compare with the variances of the VQ classifier and bagging VQ classifier.

The waveform database has 21 dimensions, and three-class simulated database[10].

It is a three-class problem based on the waveforms $h_1(t)$, $h_2(t)$, $h_3(t)$ illustrated in Figure 2.

Each class consists of a random combination of two of these waveforms sampled at the integers with noise added. More specifically, to generate a Class 1 vector $\boldsymbol{x}$, independently generate a uniform random number u and 21 random numbers $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{21}$ normally distributed with a mean of zero and a variance of one. Then set

$$x_m = uh_1(m) + (1-u)h_2(m) + \varepsilon_m,\ m = 1, \ldots, 21. \quad (8)$$

To generate a Class 2 vector, repeat the preceding step and set

$$x_m = uh_1(m) + (1-u)h_3(m) + \varepsilon_m,\ m = 1, \ldots, 21. \quad (9)$$

A Class 3 vector is generated by

$$x_m = uh_2(m) + (1-u)h_3(m) + \varepsilon_m,\ m = 1, \ldots, 21. \quad (10)$$

These 3 different vectors are generated using prior probabilities of $\{⅓, ⅓, ⅓\}$. That is, each class has the same probability. The code for generating the waveform data is in [8,11].

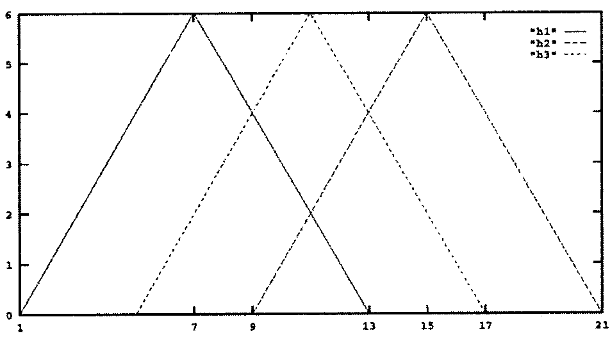

Figure 2. The $h_1(t)$, $h_2(t)$, and $h_3(t)$ functions.

We generate 100 sets of waveform database. Each set has 300 factors, so, the total amount of factors is 30,000. The factors are divided into three classes.We compute the bias and variance of the VQ classifier by Equations (4)-(7) with the waveform database. To obtain the Bayes classifier, a probability $P(y \mid \boldsymbol{x})$ is needed. The $P(y \mid \boldsymbol{x})$ is

$$P(y \mid \boldsymbol{x}) = \frac{P(\boldsymbol{x} \mid y)P(y)}{P(\boldsymbol{x})}, \quad (11)$$

We already know that the $P(y)$ is ⅓. Now, we compute the $P(\boldsymbol{x})$. Let a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_{21})$ be Class 1's vector. We will find the joint distribution $P(\boldsymbol{x})$ and each x_m can be written as follows:

$$x_m = u(h_1(m) - h_2(m)) + \varepsilon_m + h_2(m) \quad (12)$$

where $h_2(m)$ is constant, u and ε_m are random variables. Then, this problem becomes the function of 21 random variables[12].

We rewrite Equation (12) as

$$\begin{aligned} x_1 &= a_1 u + \varepsilon_1 + h_2(1) \\ x_2 &= a_2 u + \varepsilon_2 + h_2(2) \\ &\vdots \\ x_{21} &= a_{21} u + \varepsilon_1 + h_2(21) \\ y &= u, \end{aligned} \quad (13)$$

where $a_m = h_1(m) - h_2(m)$, $m = 1, \ldots, 21$.

Solve for ε_m in terms of x_m and y, then we have

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{21} \\ u \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & -a_1 \\ 0 & 1 & \cdots & -a_2 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -a_{21} \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 - h_2(1) \\ x_2 - h_2(2) \\ \vdots \\ x_{21} - h_2(21) \\ y \end{pmatrix}. \quad (14)$$

Since

$$\varepsilon_m + h_2(m) = N(h_2(m), 1). \quad (15)$$

By the transformation method, we obtain the conditional joint pdf. of x as follows:

$$P(x \mid class\,1) = \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_1 - a_1 y - h_2(1))^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2 - a_2 y - h_2(2))^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_{21} - a_{21} y - h_2(21))^2}{2}} dy \quad (16)$$

Similarly we obtain $P(x \mid class\,2)$ and $P(x \mid class\,3)$. Thus, the joint pdf. $P(x)$ of all three classes is given by

$$\begin{aligned} P(x) &= P(x \mid class\,1)P(class\,1) + P(x \mid class\,2)P(class\,2) + P(x \mid class\,3)P(class\,3) \\ &= \frac{1}{3}\int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_1 - a_1 y - h_2(1))^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2 - a_2 y - h_2(2))^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_{21} - a_{21} y - h_2(21))^2}{2}} dy \\ &+ \frac{1}{3}\int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_1 - b_1 y - h_3(1))^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2 - b_2 y - h_3(2))^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_{21} - b_{21} y - h_3(21))^2}{2}} dy \\ &+ \frac{1}{3}\int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_1 - c_1 y - h_3(1))^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_2 - c_2 y - h_3(2))^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_{21} - c_{21} y - h_3(21))^2}{2}} dy, \end{aligned} \quad (17)$$

where $a_m = h_1(m) - h_2(m)$, $b_m = h_1(m) - h_3(m)$, $c_m = h_2(m) - h_3(m)$, $m = 1, \ldots, 21$.

Now we can obtain $P(y \mid x)$ Equation (11). Therefore, the Bayes classifier C^* is obtained. Secondly, we construct the VQ classifier corresponding to each waveform database set and integrate the aggregated classifier C_A.

The variance is computed as follows: We already make the one hundred sets of 300 factors and obtained the Bayes classifier C^*, the VQ classifier C corresponding to each set, and the aggregated classifier C_A. An additional set of 18,000 factors was generated from the same distribution and the aggregated classifier was computed. According to Equation (5), this test set was divided into bias and unbias sets, and then, the bias and variance are computed by Equations (6) and (7). The bias and variance of the VQ classifier are in Table 1.

Table 1. Bias and variance with waveform database.

	CART	Bagging CART	VQ	Bagging VQ (%)
Bias	1.7	1.4	3.54	2.97
Variance	14.1	5.3	11.95	4.35
Error	29.0	19.5	24.5	13.2

Table 1 shows that the variance of the VQ classifier is as large as that of the CART.

Another experiment is also performed to show that the VQ classifier is unstable. A classifier is unstable in that minor changes in the training set could cause large changes in classifier results. We construct the bagging VQ classifier for ASR with a database-256[13] and the TIMIT database[14].

The database-256 consists of 256 utterances extracted from sixteen speakers (eight males, eight females). Each speaker uttered sixteen sentences. The average length of the utterances is about 2.36 sec. These utterances were numbered from 1 to 16. Only one sentence between 1 and

Table 2. Recognition rates using only one sentence for training (database-256).

Sentence index	Recognition rates(%)	Sentence index	Recognition rates(%)
1	21.9	6	28.1
2	24.2	7	26.6
3	25.4	8	26.6
4	28.9	9	27.8
5	28.1	10	30.9

Table 3. Recognition rates using only one sentence for training (TIMIT database).

[illegible]	[illegible]	[illegible]
1	302	20.1
2	245	24.0
3	299	18.2
4	257	16.9
5	314	19.1

10 is used by each speaker for training. The six sentences (from 11 to 16) are used for testing. As for the TIMIT database, one sentence between 1 and 5 is used by each speaker for training. The five sentences from 6 to 10 are used for testing. The experimental results are shown in Tables 2 and 3.

In Table 2, the recognition rates vary from 21% to 31%, depending on the training sentence. The recognition rates vary from 16.9% to 24% with the TIMIT database.

Therefore, according to these results, we can say that the VQ classifier is unstable. So we expect that the bagging method improve VQ classifier performance significantly.

IV. Bagging VQ classifier for ASR

Let training set L consist of data $\{(x_n, y_n), n=1,\dots,N\}$, where N is the total amount of training data, and y is the speaker's ID number. Put equal probabilities $\frac{1}{N}$ on each sample and, using these probabilities, sample with replacement N times from the training set L, forming the resampled training set L_b. Some samples in L may not appear in L_b; some may appear more than once. Use L_b to construct the corresponding classifier C^b. In the classification procedure, for an unknown feature vector x, the predicted speaker y of x is elected by Equation (2), where M is the number of resampled data sets.

The above-described algorithm is adjusted as follows:

- Training procedure:

  Step 1: Make a resampling data set L_b from the training data set L, $|L_b|=|L|$.

  Step 2: Construct VQ classifier C^b corresponding to each resampling data set L_b.

  Step 3: Repeat Steps 1 and 2 from $b=1$ to M.

After the training procedure, the resampling data set $\{L_b\}_{b=1}^{M}$ and corresponding classifiers set $\{C^b\}_{b=1}^{M}$ are obtained.

- Test procedure:

  Step 1: Perform speaker recognition by each C^b with test data t.

  The recognized speaker's ID number y_b is obtained from the classifier C^b.

  Step 2: Repeat Step 1 from $b=1$ to M.

  Step 3: Vote for the final speaker from $\{y_b\}_{b=1}^{M}$ by the majority rule.

The diagram of our proposed bagging VQ classifier for ASR is shown in Figure 3.

V. Experimental results

The ASR experiments are conducted using the bagging VQ classifier. A 100-speaker subset of the TIMIT database and the database-256 were used for the experiments. The speech features were extracted on a frame rate of 10msec

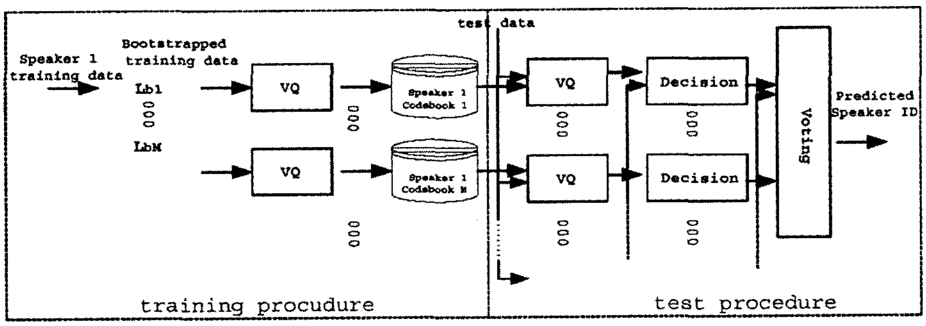

Figure 3. The bagging VQ classifier.

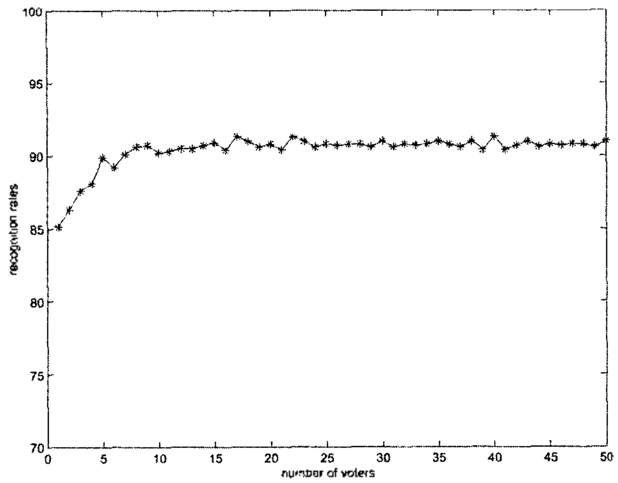

Figure 4. Recognition rates by bagging VQ classifier.

and a frame size of 30msec. The pre-emphasis (0.97) and hamming window were used. The feature vector includes 12 Mel-cepstra.

The recognition results from the conventional VQ classifier are shown in Table 4. Using the conventional method, we obtain the recognition rate of about 85%. The

Table 4. Recognition rates by conventional VQ classifier.

[illegible]	[illegible]
Codebook size	32
Recognition rates	85.0 (%)

Table 5. Recognition rates depending on number of training sentences.

[illegible]	[illegible]	[illegible]
Recognition rates	88.67 (%)	92.19 (%)
Number of training vectors	1101	2344

Table 6. Recognition rates by bagging VQ classifier (database-256, using 5 sentences for training).

[illegible]	[illegible]
2	90.63 (%)
3	94.53 (%)
4	95.70 (%)
5	95.70 (%)
6	96.48 (%)
7	97.27 (%)
8	96.48 (%)
9	96.88 (%)
10	96.88 (%)

results of the proposed bagging VQ classifier are shown in Figure 4. Each VQ codebook has 32 codewords.

Figure 4 displays the recognition rates vs. the number of voters. As shown in Figure 4, the recognition performance improves when at least two or three voters are used. The recognition rates of the conventional VQ is 85%. But the recognition rates of the bagging VQ with five voters is about 90%. The speaker recognition rates are improved significantly by bagging VQ. When the number of voters exceeds five, the performance improvement becomes marginal.

The second experiment concerns small training databases. The bagging VQ classifier reveals good performance even with small training databases. The experiments were performed with a database-256. Experimental results are shown in Tables 5 and 6.

Table 5 presents the results obtained from the conventional VQ classifier (codebook size of 16). This table shows the variation in recognition rates depending on the size of the training database. The recognition rate of 88.67% is obtained when we use 5 sentences for training. When we increase the number of training sentences to 10, the recognition rate improves to 92.19%.

The recognition results from the bagging VQ classifier (codebook size of 16) are shown in Table 6.

These classifiers use only 5 sentences for training. When at least three voters are used for the bagging, the recognition rate improves to 94.53%.

The bagging VQ classifier outperforms the conventional VQ classifier even with small training databases.

VI. Concluding remarks

In this paper, the bagging VQ classifier for ASR is proposed. We employ two experiments to show that the VQ classifier is unstable. The variance of the VQ classifier is computed and compared with that for CART. The unstability of the VQ classifier is also shown by a speaker recognition experiment. The recognition results vary significantly, depending on the construction of the training data set.

The proposed classifier is tested using the TIMIT database and the database-256. Experiments involving a closed set, text-independent and speaker identification system are done to compare the performance of the bagging VQ classifier with that of the conventional VQ classifier. Experimental results show that the bagging VQ classifier outperforms the conventional VQ classifier.

References

1. L. Breiman, "Arcing Classifiers," *The Annals of Statistics*, vol. 26, no. 3, pp. 801-849, 1998.
2. T. Matsui and S. Furui, "Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMMs," *Proc. of the ICASSP'92*, pp. II-157 - II-160, 1992.
3. D. Reynolds, *A Gaussian Mixture Modeling Approach to Text-Independent Speaker Identification*, Ph.D. Thesis of Georgia Tech., August, 1992.
4. D. K. Burton, "Text-Dependent Speaker Verification Using Vector Quantization Source Coding," *IEEE Tran. on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, No. 2, pp. 133-143, 1987.
5. R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen and V. Zue, "Survey of the State of the Art in Human Language Technology," European Communities, *http://cse.ogi.edu/CSLU/HLTsurvey/*, pp. 42-70, 1995.
6. H. S. Lee, "Speaker Recognition Technique," Proc. of the Speech Comm. and Signal Processing Workshop, pp. 42-46, 1995.
7. L. Breiman, "Bagging Predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996.
8. waveform database, *ftp://ftp.ics.uci.edu/pub/machine-learning-databases*, 1988.
9. S. H. Lee, *Tree-based Modeling of Prosody for Korean TTS Systems*, Ph.D. thesis, KAIST, 1998.
10. J. H. Friedman, R. A. Olshen, C. J. Stone, and L. Breiman, *Classification And Regression Trees*, Wadsworth Statistics /Probability Series, Belmount, CA, 1984.
11. ISTAT 5.4, *http://www.acm.org/~perlman/stat*, 1998.
12. A. Papulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.
13. Y. J. Kyung and H. S. Lee, "Text-independent speaker identification using micro-prosody," *Proc. of ICSLP '98*, vol. 2, pp. 157-160, 1998.
14. TIMIT, *DARPA TIMIT-Acoustic-Phonetic Continuous Speech Corpus*, NIST, 1993.

[Profile]

• YounJeong Kyung

YounJeong Kyung received the B.S. and M.S. degrees in computer science from Dongduk Women"s University, Seoul, Korea, in 1992 and 1994. She received the Ph.D. degree in electrical engineering and computer science (EECS) from KAIST, Daejon, Korea, in 2000. Since 2000, she has been with SK Telecom as a researcher. Her current research interests include VXML based voice service platform and speaker recognition on wireless network.

• BongDae Choi

BongDae Choi received the B.S. and M.S degrees in mathematics from Kyungpook National University, Daegu, Korea in 1970 and 1972. He received the Ph.D. degree in Mathematics from Ohio State University, U.S.A, in 1980. He was a professor at the KAIST from 1983 to 1998. Since 1998, he has been with the department of Mathematics, Korea Univ., Korea, as a professor. His current research interests include applied probability.

• HwangSoo Lee

HwangSoo Lee received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea in 1975. He received the M.S. degree and the Ph.D. degree in electrical engineering from KAIST, Seoul, Korea, in 1978 and 1983. Since 1983, he has been with the department of electrical engineering and computer science, KAIST, Korea, as a professor. He also has joined with SK Telecom as a advisor since 1997. His current research interests include All IP network, wireless network and speech signal processing.