Recognition of Speech Isolated Words Based on Pyramid Phonetic Bag of Words Model Display and Kernel-Based Support Vector Machine Classifier Model

Sodabeh Salehi Rekavandi, Hamidreza Ghaffary and Maryam Davodpour

Abstract This study aimed to improve the classification of individual (isolated) words, and specifically, the numbers from one to twenty. In this study, a strong model was suggested to gain a unified view of voice. It is based on the idea of phonetic bag for voice that has been developed into a pyramid state. The pyramid idea can model temporal relationships. One of the problems of Support Vector Machine to classify words is its inability to model temporal relationships unlike hidden Markov models. Using the BOW-based pyramid idea in the extraction of the display containing temporal information of voice, the SVM can be given the capability of considering the time relationships of speech frames. One of the main advantages of Support Vector Machine model is its fewer parameters than the hidden Markov model. As the experiments' results have shown, it has much higher accuracy than the hidden Markov model in applications such as the recognition of single words, where the data set volume is limited. Using the pyramid BOW idea, the accuracy of SVM-based method can be increased as 20% compared to previous methods.

Keywords Speech recognition · Isolated words recognition · Classification of speech introduction · Display of phonetic bag of words · Support vector machine

S. S. Rekavandi (✉) · H. Ghaffary · M. Davodpour
Department of Computer Engineering, Islamic Azad University, Ferdows, Iran
e-mail: s_salehi19@yahoo.com

H. Ghaffary
e-mail: hamidghaffary53@yahoo.com

M. Davodpour
e-mail: Maryam.Davodpour2@gmail.com

S. Montaser Kouhsari (ed.), *Fundamental Research in Electrical Engineering*, Lecture Notes in Electrical Engineering 480,
https://doi.org/10.1007/978-981-10-8672-4_2

1 Introduction

In this study, an efficient method based on pyramid bag of words (BOW) model and the SVM classifier model were provided to recognize isolated words. The provided BOW method has the ability to describe and model the temporal relationships in the speech, and by using kernel-based nonlinear support vector machine model can be used as an efficient technique used in recognition applications of isolated words.

Hedges et al. [1] studied the isolated words recognitions word using the support vector machine. In this method, first, the voice framed, and the Mel Frequency Cepstral Coefficients (MFCC) features extracted from each frame.

This stage is common in the most speech processing studies, and it indeed models a descriptive frequency of the frame. In fact, we expect the corresponding frames to have a MFCC feature vector similar to a particular part of a phoneme (e.g., frames related to explosion part of the explosive phoneme "b"). In other words, the difference is expected to be negligible. In this study, this stage as a conventional tool in describing a frame is constant in all discussing suggested methods. In their approach [1], the MFCC characteristics of each frame of a word (sound) is given to the Support Vector Machine (SVM) Classifier with the label of that word. For example, suppose a sound with the tag of "Five" includes 100 frames in 32 ms (with taking into account the overlap). Of these 100-frame, we calculate100 MFCC feature vector. Each of these 100 vectors (39-next) are labeled as "Five" and insert into the classifier. The same process is repeated during testing the training model with these difference that 100 labels predict by the SVM model. To obtain the label, the majority vote is considered among the 100 obtained predictions. This strategy has two major problems which we resolve them in this study.

To understand the first problem, consider this example that the phoneme "I" exists in both words of "Five" and "Nine". Thus, this method gives the frames related to this phoneme to the classifier with two different labels. Regardless of the classifier model, this strategy will disrupt the learning process of the model. In this research, we have resolved this problem by generating a unified display of speech based on bag of word (BOW) techniques. The second problem is the lack of modeling of temporal relationships in recognizing the words. In this study, using the pyramid-making idea of displaying BOW (Pyramid BOW), which has been highly regarded in recent years in the processing of images for modeling the spatial relationships, we provide a pyramid display model for voice (sound) that can model the temporal relationships (transposition of frame).

Models such as hidden Markov model inherently model the temporal relationships in the sound. However, in this study, we have used support vector machine as the classifier model.

The disadvantage of HMM models is their failure to have sufficient efficiency in small applications and recognizing isolated words. As a result, we would require massive datasets for their training. In fact, the number of HMM model parameters is very high, and in order to prevent the model overfitting, we need a lot of data. In HMM model, we need only to train a HMM model per word with a sufficient

number of modes (for example, 6 modes). In each of these modes, we need to estimate the conditional probability of all observations. Suppose that the observations are possible for 50 MFCC models. Each of these patterns is related to different passes of one of the phonemes (e.g., the explosive section of "B").

Thus, we need to estimate 6 * 50 conditional probabilities for each word. For 20 words, this number is 6000 parameters, which is a large figure compared to the number of data. However, the parameters can be somewhat reduced by techniques such as modeling at the phoneme level (each HMM models a phoneme). Of course, using such techniques requires providing the label at the level of the phonemes, which is a very time-consuming process; and at the same time, even if we consider two states for each phoneme, we should estimate 100 parameters, and to estimate the probabilities, we should have a high number of phonemes which do not practically make a significant change in the applications such as recognizing isolated words, but it can be used for continuous speech recognition. In methods like Support Vector Machine, using techniques such as reducing dimension, the number of model parameters can be controlled, and the overfitting of model can be prevented. Thus, the dimensionality reduction technique of principal component analysis (PCA) is raised. Therefore, this method is used to reduce the BOW-based feature vectors.

The results show the effectiveness of proposed methods to classify the isolated words.

2 Prior Research

In this section, first, the stages of extracting common characteristics of the sound signal are described. Then, the background of works related to displaying the bag of words and classification are described. In the next section, this method has been developed to classify speech isolated words.

3 Pre-processing and Feature Extraction

Several stages of recognition system are performed in the preprocessing phase. First, the speech is segmented into frames. Usually in speaker recognition applications, for better performance, the noise parts and the speech silence are eliminated. In this study, we have applied this stage as well.

In all branches of speech processing (speech recognition, word finding, speaker identification, etc.), the second phase is to extract feature from speech frames. Different feature vectors have been used for speech, including linear prediction coefficients, Mel Frequency Cepstral Coefficients (MFCC), wavelet coefficients and so on. In this study, the best and most effective ones, the MFCC has been used.

The Mel Frequency Cepstral Coefficients (MFCC) have been known as the most common and most widely used feature vector in processing of voice (audio) signal. After obtaining the filters bank energy, the feature vector of MFCC will be achieved by using discrete sine-cosine transform. In this section, the stages of feature extraction are explained below. The output of this stage is a feature vector sequence that each has been extracted from one of the input speech frames.

3.1 First Stage: Removing Silence from the Beginning and End of Words

In this research, for better efficiency, the silence at the beginning and end of words has been deleted using the method presented in [2]. This method has been implemented in MATLAB software at high speed[1]. This implementation is used in this study. The output of this method includes segments containing speech activity that the word's part of speech can be achieved by incorporating them. Voice Activity Detection (VAD), which is also called speech activity detection or speech recognition, is a process in the area of speech processing in which the presence or absence of human speech is recognized. Although the main use of this technique is in speech encoding and speech recognition, but it is also used in some other activities, such as speaker recognition. The goal in this method is to separate speech parts from silence and non-speech parts. The voice active areas usually refer to areas that are not related to environmental noise or silence. VAD methods extract parameters such as Linear Predictive Coding (LPC) distance, energy and zero crossing rate and compare these parameters with a set of threshold values to detect intervals including speech. Since these threshold values are estimated by analysis of silence periods, the classification accuracy of these methods highly reduces under unfavorable acoustic conditions. Normally, there is only noise in areas of the signal with silence. Through this measure with the ability to detect pure noise, it is possible to detect silence in the signal. The VAD problem is usually challenging in terms of low signal-to-noise (low SNR). Low SNR along with unstable noise signal can greatly reduce the precision of a VAD system. The basic methods for VAD detecting are based on signal energy. However, this measure does not work well when the SNR is low, since the energy of parts with sound activity is almost identical to noisy areas, and even in the unstable noise of energy, a measure becomes quite useless. The algorithm of method used to remove the silence at [2] is fully described.

[1] http://www.mathworks.com/matlabcentral/fileexchange/28826-silence-removal-in-speech-signals/

3.2 Second Stage: MFCC Feature Extraction Method

In this section, the detailed steps of MFCC method used in this study are described.

Suppose that $s_1, \ldots, s_{512}$ are examples of the studied frame. The stages of MFCC method used for each frame are as follows:

- Frame energy calculation: The mean square of frame samples

$$E = \frac{\sum_{i=1}^{512} s_i^2}{512}. \tag{1}$$

Applying 512-point Hamming window on $s_1, \ldots, s_{512}$

$$s_{w-1}, \ldots, s_{w-512} = h_1 s_1, \ldots, h_{512} s_{512} \tag{2}$$

- Calculating the FFT of windowed frame

$$f_1, \ldots, f_{512} = fft(s_{w-1}, \ldots, s_{w-512}) \tag{3}$$

- Calculating the result of 12 Mel filters (12 channels) on $f_1, \ldots, f_{512}$. At this stage, 12 $Z_1, \ldots, Z_{12}$ are obtained.
- Calculating 13 features by using the following equation:

$$\begin{aligned} mfcc_1, \ldots, mfcc_{12} &= DCT(\log(Z_1, \ldots, Z_{12})) \\ mfcc_{13} &= \log(E) \end{aligned} \tag{4}$$

Calculating 13 features by using the derivative of $mfcc_1, \ldots, mfcc_{13}$:

These features are called Delta. To calculate the derivative, every two consecutive numbers are subtracted (The first number is subtracted from the last number). The feature obtained in this step are called as $d_1, \ldots, d_{13}$.

- Calculating 13 features by using the derivative of $d_1, \ldots, d_{13}$: These features are called Delta-Delta. At this stage, the $dd_1, \ldots, dd_{13}$ is obtained.
- The final feature vector is as follows (including 39 real number):

$$F = [mfcc_1, \ldots, mfcc_{13}, d_1, \ldots, d_{13}, dd_1, \ldots, dd_{13}]. \tag{5}$$

Therefore, for each studied voice frame, a 39-item feature vector is extracted.

4 Display of Bag of Words

The display of bag of words (BOW) has been primarily inspired in the field of image processing from the field of text processing [3]. As the number of each word can be easily counted within a text, our goal here is to count the patterns in an image or a sound. In using BOW-based methods in image, initially, the possible patterns in a dictionary are learned. For example, an eye pattern can be one of the patterns available in the dictionary. This idea has been widely used in image processing [4–6]. In audio processing tasks, this method has been sometimes introduced as Bag of Acoustics [7]. This method has been regarded in recent years in the issue of sense detection [7] and recognition voice from event [8].

5 Dimensionality Reduction of Principal Component Analysis

In dimensionality reduction methods, a multi-dimensional space are mapped to a space of lower dimension. With reducing the dimensions of the original space, the number of model parameters would reduce, and thus, the probability of model overfitting will decrease. PCA dimensionality reduction is as such to maintain information as much as possible. In addition to this feature, the PCA method finds the direction of highest changes and depict the data in those directions. Therefore, it is a useful feature transfer method that is used in most applications of pattern recognition [9–11]. In this study, after extraction simple and pyramid BOW display provided, this method has been used to reduce the dimensions.

6 Suggested Method

In this section, first, the proposed method for finding a BOW-based display of input speech is described. Then, the idea has been developed to model temporal relationships in the speech into a pyramid way. Finally, the diagram block of the proposed method is provided.

7 Display of Bag of Acoustics

Figure 1 shows the proposed method to obtain a BOW display of an acoustic signal.

As can be seen in Fig. 1, a dictionary including K patterns (templates) is provided (The dictionary learning method is described in the next section). Each input

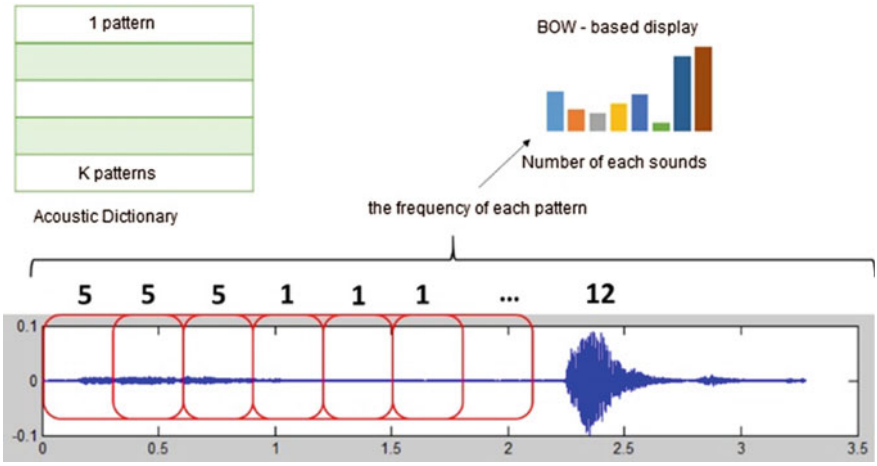

Fig. 1 Extraction of BOW display from a sound (first suggested method)

sound is divided into consecutive frames with overlapping. The MFC features are extracted from each of these frames. For each frame, the closest MFCC model in the dictionary is found. After this stage, the number of each model can be counted. Therefore, we have a display resulting from the frequency of K patterns in the sound. In this study, we will solve one of the fundamental problems of the basic method by using the BOW method, in which each frame is given to the classifier separately. However, there is another problem in this method. Although we have considered many different acoustic patterns in the display of sound, but no information has been modeled about their order. This problem has been solved by using the idea of pyramid-making of BOW display [12], which has been highly regarded in recent years in images processing for modeling spatial relationships [13, 14].

8 Learning of Phonetic (Acoustic) Dictionary

To learn phonetic (acoustic), dictionary any clustering method can be used. In this study, we have used the known k-means method. First, the 39-item MFCC vectors are extracted from all frames of total sounds in the training data set. The goal is to learn K cluster centers (phonetic pattern) of these vectors in such a way that the quantization error is so small. Quantization error refers to the difference of each vector with the nearest cluster, i.e. a cluster that belongs to it.

Therefore, at this point, it is assumed that M MFCC vectors have been selected as $S = \{s_1, s_2, \ldots, s_M\}$. Now, it is just enough to train K patterns of the vectors within the S. to this end, the S vectors must be grouped into K clusters $\{S_k, k = 1, 2, \ldots, K\}$, while clusters have patterns different from each other. For this

Algorithm 1: Clustering

Inputs: all MFCC vectors in S, K
Outputs: $\mu_1, \ldots, \mu_K$.

Step 1: Intialize $\mu_1, \ldots, \mu_K$
for $k = 1$ to K do
$\mu_k \leftarrow random\ sample\ from\ S.$
end for
Step 2: Assign and Update Iteratively
while max iterations do
for $i = 1$ to M do
Assign s_i to Cluster that minimize $||S_i - \mu_k||_2^2$.
end for

for $k = 1$ to K do
Update μ_k Using Mean of $\{s_i | s_i \in c_k\}$.
end for
Step 3: Remove Useless Clusters
Every Cluster with no member removed.

Fig. 2 Clustering algorithm to learn the phonetic dictionary

reason, it is enough to do the clustering based on MFCC features, since it is proportional to the human auditory system.

If the k-means clustering algorithm is applied on these vectors, the vectors in the S are divided into K clusters, $C_1, \ldots, C_K$, and the μ_k phrase is chosen as the center of cluster C_k. *The* K-means clustering algorithm is as follows: (Fig. 2).

In the first phase of algorithm 1, the centers are initialized. The second phase varies repeatedly between the two stages of attributing to the cluster centers and updating the centers until reaching the desired number of repetitions. The third step removes all clusters with no members. After applying the k-means algorithm, the cluster centers show the intended phonetic dictionary in the MFCC space.

9 Display of Pyramid Bow to Model Temporal Relationships in Speech

The pyramid-making idea to fix the problem of BOW display in modeling spatial relations in the image was raised for the first time in [9], and has been of great concern in the field of image processing so far. Models such as hidden Markov models inherently model temporal relationships in the sound. But as noted, in this study, we have used the support vector machine as the classifier model.

The disadvantage of HMM models, which causes their inefficient use in small applications and recognizing individual words, is the need to massive dataset for training them. In fact, the number of HMM model parameters is very high, and a lot of data is required to prevent the model overfitting. In methods such as support vector machines, using techniques such as dimensionality reduction, the number of model parameters can be controlled, and the model overfitting can be prevented. The idea of pyramid-making of BOW relies on fragmentation of the image to the required level and calculating the frequency of patterns in each slice (Fig. 3). For example, if we tell someone that there are two models of eyes and a nose pattern in an image, it cannot be expected that the person can guess where on the image the patterns occur. But with pyramid-making of BOW display, this problem goes away.

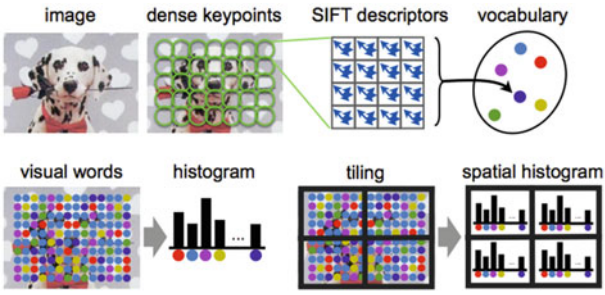

Fig. 3 BOW pyramid display in the image

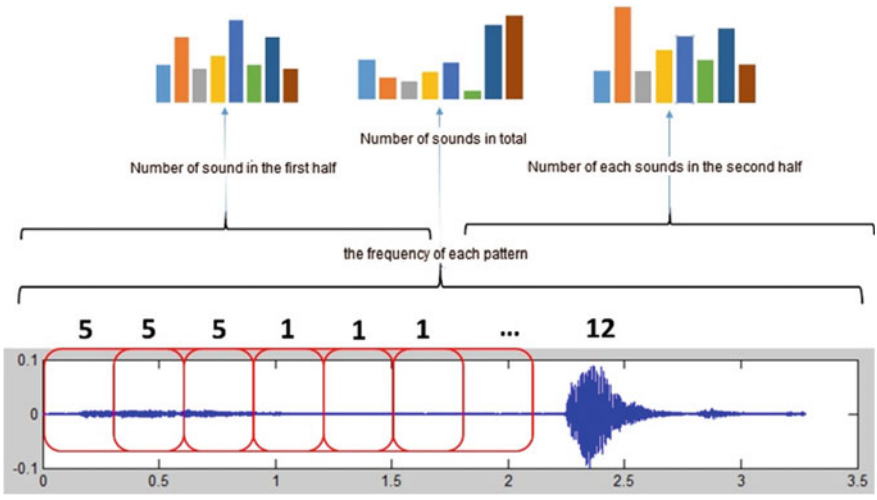

Fig. 4 Pyramid display of BOW in the sound (second alternative method)

In this study, we have used the idea of pyramid-making for modeling temporal relationships in the sound. Figure 4 shows the proposed approach for pyramid making of display in the sound.

Two levels are used in Fig. 4. If needed, the number of levels can be increased. In the next level, four areas are achieved. If the words are long and the number of phonemes of each word is high, higher number of levels would more appropriate.

10 Diagram Block of the Proposed Method

In previous sections, the proposed methods based on simple and pyramid BOW display were described. In this section, the steps of the proposed method are summarized. Figure 5 shows the diagram block of the model teaching stage.

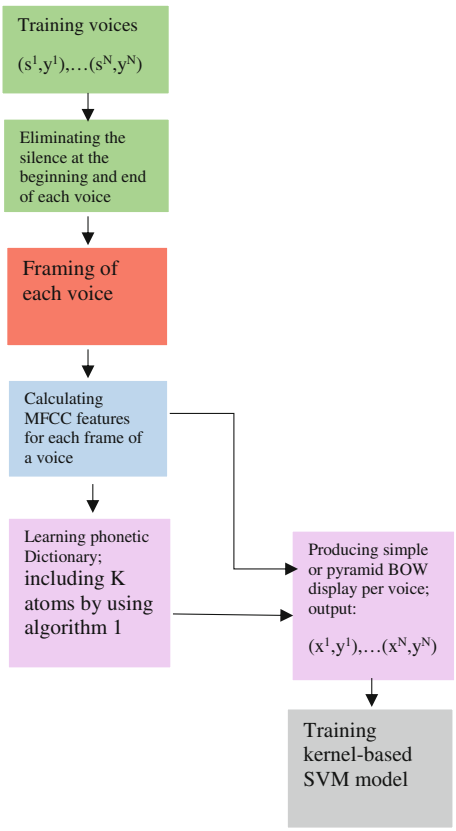

Fig. 5 Diagram block of learning algorithm of the classifier model

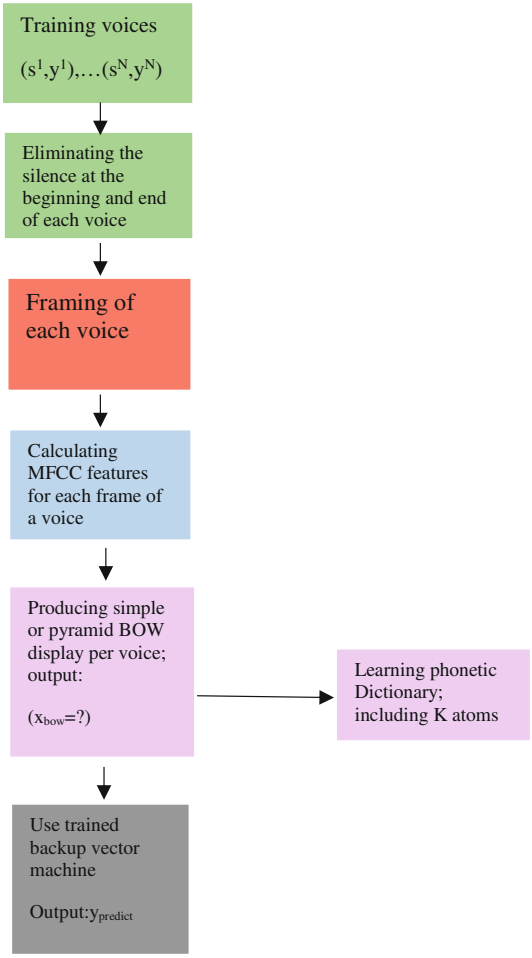

Fig. 6 Block diagram of classifying a new data

After teaching the phonetic dictionary and the classifier model, they can be used to classify a new data. The diagram block is the use of the trained model to predict the label of a data as Fig. 6.

- Training voices
- Eliminating the silence at the beginning and end of each voice
- Framing of each voice
- Calculating MFCC features for each frame of a voice
- Learning phonetic Dictionary; including K atoms by using algorithm 1
- Producing simple or pyramid BOW display per voice; output

- Training kernel-based SVM model
- Training voices
- Eliminating the silence at the beginning and end of each voice
- Framing of each voice
- Calculating MFCC features for each frame of a voice
- Learning phonetic Dictionary; including K atoms by using algorithm 1
- Producing simple or pyramid BOW display per voice; output
- Training Kernel-based SVM model.

11 Experiments

In this section, we test the basic techniques and the described method. First, the training data set is described. Then, the evaluation criteria are described. Finally, the settings related to experiments and the test results are given.

12 Training Data Set

Training data set includes 10 different speakers. Each of these speakers have uttered words 1–20 once, the words with sampling frequency of 16 kHz have been recorded. The data related to 7 speakers have been selected as training data, and 3 speakers as the test data.

13 Noisy Dataset

A noisy data set has been also made to evaluate the effectiveness of models in the presence of ambient noise. This data set has four samples per voice (in the previous section):

- Original sound version
- Signal-to-noise: 30 dB
- Signal-to-noise: 20 dB
- Signal-to-noise: 10 dB

Therefore, this dataset contains 28 samples per word in the training data set and 12 samples per word in the test data set.

14 Feature Extraction

The feature is segmented to 32 ms (ms) frames with an overlap of 16 ms. For better performance, as mentioned in previous sections, the noise and silence parts have been removed before framing. At this stage, by the number of frames per voice, the MFCC 39-fold vectors are obtained.

15 Classification Tests Results

Table 1 shows the test results of basic models and suggested methods. The HMM method is indeed discrete HMM method. The number of optimal HMM states for words was obtained as 6. The number of clusters in the HMM method was equal to 40 optimized ones, while the number of clusters in BOW-based and pyramid BOW methods was equal to 100 optimized clusters. In BOW and pyramid BOW methods, the PCA method was used to reduce the dimensions.

16 Classification Analysis

16.1 Number of Optimal Clusters

The results of Table 1 show that HMM method has low accuracy in the recognition of isolated words, and it was expected due to the high number of HMM model parameters compared to the number of data. The number of optimal clusters in HMM method (40 clusters) was lower compared to methods based on SVM (100 clusters). A total of 40 clusters, especially in noisy data, cannot model a variety of phonetic patterns. However, by increasing the number of clusters instead of increasing the precision, we would have accuracy reduction as well. The reason for this phenomenon is that by increasing the number of clusters, the number of model parameters will extremely increase and there is no way to control the number of

Table 1 Classification accuracy of isolated words

Method name	Accuracy of the normal data set		Accuracy of the noisy data set	
Words	20-1	5-1	20-1	5-1
Random	5%	20%	5%	20%
HMM	28%	53.3%	15.3%	41.6%
Basic method [1]	31%	59.4%	18.2%	46.6%
Display + BOW SVM	*50.3%*	**73.3%**	*45.8%*	**63.3%**
Display BOW + Pyramid SVM	*72%*	**96%**	*68.6%*	**84.6%**

parameters. The number of parameters of SVM model is inherently lower than the HMM method. In addition, the use of PCA method can control the number of parameters.

17 Comparing the SVM-Based Method with Bow SVM

The results show that the proposed method in [1, 15], which has been reported as the basic method in Table 1, has also a less accuracy than the proposed conventional BOW method. In basic method [1], the MFCC features of each frame from every word (sound) are given tagged with that word to the support vector machine classifier. For example, suppose a sound with the tag of "Five" includes 100 frames in 32 ms (with taking into account the overlap). Of these 100 frames, 100 MFCC feature vectors are achieved. Each of these obtained 100 vectors (39-next) are labeled "Five" and given to the classifier. The same process is repeated when testing the training model; the difference is that 100 labels are predicted by the SVM model. To obtain the label, the majority vote is taken among the 100 predictions. This strategy has two major problems, which are addressed in this study. To understand the first problem, consider this example that the phoneme "I" exists in both words of "Five" and "Nine". Thus, this method gives the frames related to this phoneme to the classifier with two different labels. Regardless of the classifier model, this strategy will disrupt the learning process of the model.

18 Comparison of Bow Method with Pyramid Bow Approach

The results show a significant increase in the accuracy of the pyramid display compared to the typical BOW. As described, the pyramid display model the temporal information in the sound and extracts more information from the sound.

19 Methods Resistance Against Noise

Among the methods, BOW and BOW pyramid methods, which are the proposed methods, have a high resistance to noise. The reason is that the noise in the MFCC features partly disappears in the quantization phase (in clustering) as clustering error, but in the basic method [1], this noise gives itself as a part of the feature vector to the SVM model. Clustering methods inherently eliminate the noise partly as quantization error.

20 Conclusion

This paper aimed to classify the isolated speech words. First, a pyramid model proposed based on BOW to display the voice which is able to increase the prediction power of support vector machine model. This model could model the temporal relationships in the sound. Using this model, the accuracy of support vector machine classifier based methods significantly improved. In this study, we demonstrated that the dimension reduction techniques are useful for increasing the accuracy of Support Vector Machine.

References

1. Hegde S, Achary KK, Shetty S (2012) Isolated word recognition for Kannada language using support vector machine. In: Wireless networks and computational intelligence. Springer, Berlin Heidelberg, pp 262–269
2. Giannakopoulos T (2009) A method for silence removal and segmentation of speech signals, implemented in Matlab. Department of Informatics and Telecommunications, University of Athens, Greece, Computational Intelligence Laboratory (CIL), Insititute of Informatics and Telecommunications (IIT), NCSR DEMOKRITOS, Greece
3. Gabriella C, Dance C, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. In: Workshop on statistical learning in computer vision, ECCV, vol 1, no 1–22, pp 1–2
4. Yang, Jun, Yu-Gang Jiang, Alexander G. Hauptmann, and Chong-Wah Ngo. "Evaluating bag-of-visual-words representations in scene classification." In: *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pp. 197–206. ACM, 2007
5. Ramesh B, Xiang C, Lee TH (2015) Shape classification using invariant features and contextual information in the bag-of-words model. Pattern Recogn 48(3):894–906
6. Yang Y-B, Zhu Q-H, Mao X-J, Pan L-Y (2015) Visual feature coding for image classification integrating dictionary structure. Pattern Recogn
7. Pokorny FB, Graf F, Pernkopf F, Schuller BW (2015) Detection of negative emotions in speech signals using bags-of-audio-words. In: 2015 International conference on affective computing and intelligent interaction (ACII). IEEE, pp 879–884
8. Grzeszick R, Plinge A, Fink GA (2015) Temporal acoustic words for online acoustic event detection. In: Pattern recognition. Springer International Publishing, pp 142–153
9. Wu P, Hoi SCH, Xia H, Zhao P, Wang D, Miao C (2013) Online multimodal deep similarity learning with application to image retrieval. In: Proceedings of the 21st ACM international conference on Multimedia. ACM, pp 153–162
10. Quan C, Wan D, Zhang B, Ren F (2013) Reduce the dimensions of emotional features by principal component analysis for speech emotion recognition. In: 2013 IEEE/SICE International symposium on system integration (SII). IEEE, pp 222–226
11. Chiou B-C, Chen C-P (2013) Feature space dimension reduction in speech emotion recognition using support vector machine. In: Signal and information processing association annual summit and conference (APSIPA), 2013 Asia-Pacific. IEEE, pp 1–6
12. Lazebnik S, Schmid C, Ponce J (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE computer society conference on computer vision and pattern recognition, vol 2. IEEE, pp 2169–2178

13. Kim SY, Sohn K-A (2015) Mobile phone spam image detection based on graph partitioning with pyramid histogram of visual words image descriptor. In: 2015 IEEE/ACIS 14th international conference on computer and information science (ICIS). IEEE, pp 209–214
14. Lan Z, Hauptmann AG (2015) Beyond spatial pyramid matching: space-time extended descriptor for action recognition. *arXiv preprint* arXiv:1510.04565
15. Seryasat OR, Aliyari-shoorehdeli M, Honarvar F (2010) Multi-fault diagnosis of ball bearing based on features extracted from time-domain and multi-class support vector machine (MSVM). In: IEEE international conference on systems man and cybernetics (SMC), pp 4300–4303