A Voice Controlled Service Robot Using Support Vector Machine

Seong Rock Kim, Jae Suk Park, Ju Hyun Park, and Suk Gyu Lee
The Department of Electrical Engineering, Yeungnam University, Gyongsan Gyongbuk, Republic of Korea
(Tel : +82-53-810-2487; E-mail: nongddiest@yumail.ac.kr,{sglee, jessie}@yu.ac.kr)

Abstract: This paper proposes a SVM(Support Vector Machine) training algorithm to control a service robot with voice command. The service robot with a stereo vision system and dual manipulators of four degrees of freedom implements a User-Dependent Voice Control System. The training of SVM algorithm that is one of the statistical learning theories leads to a QP(quadratic programming) problem. In this paper, we present an efficient SVM speech recognition scheme especially based on less learning data comparing with conventional approaches. SVM discriminator decides rejection or acceptance of user's extracted voice features by the MFCC(Mel Frequency Cepstrum Coefficient). Among several SVM kernels, the exponential RBF function gives the best classification and the accurate user recognition. The numerical simulation and the experiment verified the usefulness of the proposed algorithm.

Keywords: Support Vector Machine, Speech Recognition, Classification, kernel, Service Robot.

1. INTRODUCTION

With the development of electronic industry, robot system technology has been expanded increasingly. As one such application, the service robot with automatic voice recognition capability has been developed recently. Voice recognition can be categorized as isolated word, connected word, continuous speech, and keyword. We deal with isolated word scheme, which is easy to detect the beginning and the termination of the voice.

SVM can be used to solve the problem of pattern recognition. SVM provides the optimal hyperplane which separates a set of positive examples from a set of negative ones with the maximum margin. SVM training finds a global minimum. It is especially suitable for the learning task with the finite amount of training data given. The optimizing algorithm minimizes the classification error based on the VC theory that does not have the strong dependence of input space dimensionality. Recently, SVM has applied to various fields successfully not only in classifications but also in regression problems.[1]

SVM speech recognition algorithm has an advantage over Hidden Markov Modeling(HMM) especially when the learning data for statistical modeling are not enough.

In this paper, we propose the leaning SVM method to classify the voice vectors. Applying SVMs to Automatic Speech Recognition (ASR)[2] robot system, we have achieved good classification and the user identification.

2. Support Vector Machine Classification

Support Vector Machines proposed by Vapnik[2] decides the decision hyperplane to find the "optimal" boundary with the maximum margin between two class sets in a vector space. Each training vector x_i is classified into the class

$$(x_1, y_1), \cdots, (x_n, y_m), x \in R^n,\ y_m \in \{-1, 1\} \tag{1}$$

with a hyperplane

$$\mathbf{w} \cdot \mathbf{x} + b = 0,\quad \mathbf{w} \in R^n,\quad b \in \mathbf{R} \tag{2}$$

where **w** is the weight vector, and **b** the bias. Suppose that all the training data satisfy the following constraints,

$$\begin{aligned} x_i \cdot w + b &\geq +1 \text{ for } y_i = +1 \\ x_i \cdot w + b &\leq -1 \text{ for } y_i = -1 \end{aligned} \tag{3}$$

Then these can be combined into one set of inequalities.[3]

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \tag{4}$$

In this case, the maximum margin is given by

$$\rho(w, b) = \frac{2}{\|w\|} \tag{5}$$

To find **w**, the following cost function should be minimized in the Convex Optimization Problem of,

$$\min \ \frac{1}{2}\|w\|^2,\ \textit{subject to}\ \ y_i(wx_i + b) \geq 1,\ \forall i \tag{6}$$

So the optimization problem is to solve the following Lagrangian functional

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} \alpha_i [y_i(wx_i + b) - 1] \tag{7}$$

where α_i are the Lagrange multipliers. The Lagrangian has to be minimized with respect to **w**, b and maximized with respect to $\alpha_i \geq 0$.

This optimization problem can be converted into a dual problem of (8). To calculate **w**, b the Lagrange multipliers α_i should have maximum values in (8).

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_{i=1}^{l} \alpha_i \tag{8}$$

with constraints,

$$\begin{aligned} &0 \leq \alpha_i \leq C,\ i = 1, \ldots, l \\ &\sum_{i=1}^{l} \alpha_i y_i = 0 \end{aligned} \tag{9}$$

At boundary, there are some bound support vectors whose Lagrange multipliers equal the C parameter. The uncertain part of Cortes's approach is that the coefficient C has to be determined.[4][5]

From the accompanying KKT(Krush-Kuhn-Tucker) Conditions α_i can be solved easily.[3]

$$\alpha_i (y_i (w \cdot x_i + b) - 1) = 0 \tag{10}$$

The optimal separating hyperplane is given by,

$$\begin{aligned} \overline{w} &= \sum_{i=1}^{l} \alpha_i x_i y_i \\ \overline{b} &= -\frac{1}{2}\overline{w}[x_r + x_s] \end{aligned} \tag{11}$$

where x_r and x_s are support vectors from each class satisfying, $\overline{\alpha}_r, \overline{\alpha}_s > 0$, $y_r=1$, $y_s=-1$ respectively.

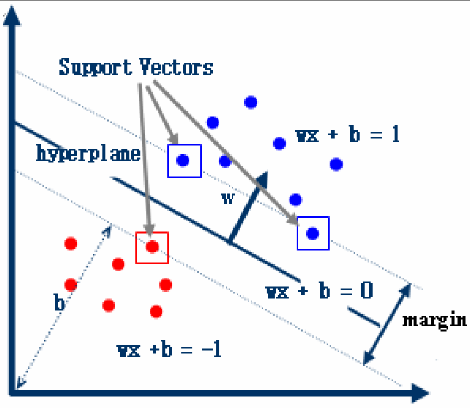

Fig.1. Separable classification linear SVMs

Each class is determined from the following discrimination function

$$f(x)_s = \text{sign}(\bar{w}\cdot x + \bar{b}) \quad (12)$$

where $f(x) = \bar{w}\cdot x + \bar{b}$

The kernel function helps deforming the classification boundary by increasing the dimension of vector space and results in the truly non-linear boundaries classification. When the kernel contains a bias term [3], we get

$$f(x) = \sum_{SVs} \bar{\alpha}_i y_i K(x_i, x)$$

$$f(x)_s = \text{sign}\left(\sum_{SVs} \bar{\alpha}_i y_i K(x_i, x)\right) \quad (13)$$

where $$\bar{w}\cdot x = \sum_{SVs} \bar{\alpha}_i y_i K(x_i, x) \quad (14)$$

Some of such kernel functions are shown below:

Linear kernel : $K(x_i, x) = x^T x_i$

Polynomial kernel: $K(x_i, x) = (x^T x_i + 1)^d$, d=1, …

Gaussian RBF: $$K(x_i, x) = \exp\left(-\frac{(x^T - x_i)^2}{2\sigma^2}\right)$$

Exponential RBF: $$K(x_i, x) = \exp\left(-\frac{|x^T - x_i|}{2\sigma^2}\right)$$

We applied these classification functions to the speech input (voice command) case of robot system. And hence our goal is to find $f(x)_s$ for each kernel functions to compare the classification performance.

3. Service Robot

The overall system structure is given in Fig 2.

The implemented service robot has CCD stereo vision system, dual manipulator, end effector(or gripper) and wheel drives. The five control boards equipped with TI(Texas Instrument) DSP TMS320LF2407A drive DC motors. The host PC calculates the algorithm and controls the manipulator. For the communication between driver controls at manipulator and vision section, Controller Area Network (CAN) is employed to improve the communication speed and robustness to the noise.[9]

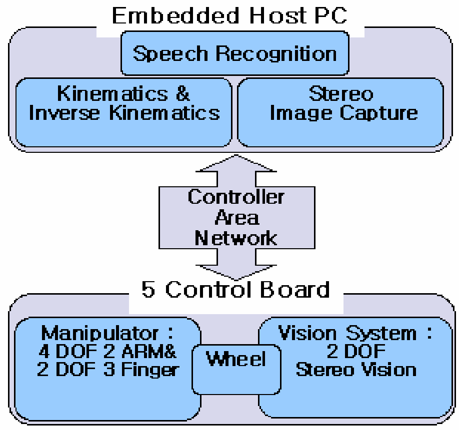

Fig 2. Robot System

4. System Structure

Each incoming speech input is converted into the feature space and this feature is compared with the SVM-trained database.

SVM discrimination function $f(x)_s$ decides the class of user's extracted feature by Mel Frequency Cepstrum Coefficient(MFCC). And then the host PC transmits the identified voice command to control Boards using CAN or serial communication.

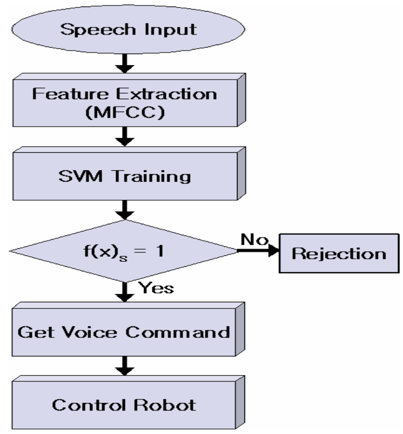

Fig.3. Structure of the proposed SVM algorithm

5. Experiments and Results

Three SVM learning results adopting three different kernels have been compared. The training data are prepared from 10 syllables repeated 10 times by one speaker. The evaluation data consists of 10 syllables repeated 3 times by 20 different speakers in addition to the training data(total 700 data).

Korean syllables used : /il/, /i-/, /sam/, /sa/,/o-/, /yook/, /chil/, /pâl/, /goo/, /sip/.

By the experiment, shown in Fig. 6, the exponential RBF kernel shows the best identification results. The classification accuracy for each kernel method is summarized in Table 1.

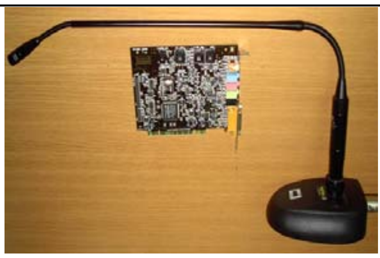

Fig. 4. The equipment of Speech Recognition

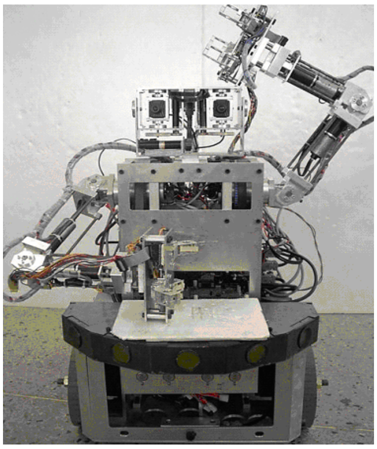

Fig.5. The Service Robot used

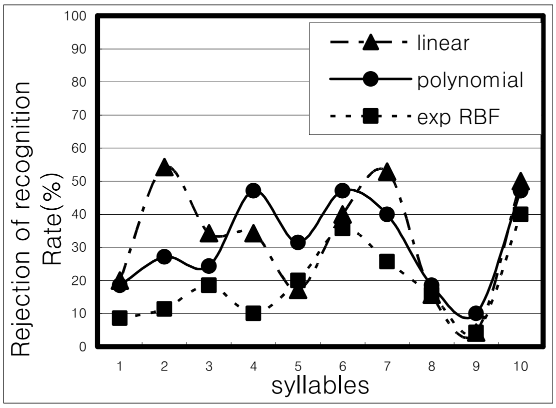

Fig 6. Rejection of recognition of speech input

The parameters of each kernel are as follows:
Linear kernel : C=10
Polynomial kernel : C=10, σ =2
Exponential RBF kernel : C=10, σ =2

Table. 1. The results of SVM by kernel method

syllables	Error Rate (%)		
(Korean)	linear	poly	exp RBF
1/il/	20.00	18.57	8.57
2/i-/	54.29	27.14	11.43
3/sam/	34.29	24.29	18.57
4/sa/	34.29	47.14	10.00
5/o-/	17.14	31.43	20.00
6/yook/	40.00	47.14	35.71
7/chil/	52.86	40.00	25.71
8/pal/	15.71	18.57	15.71
9/goo/	4.29	10.00	4.29
10/sip/	50.00	47.14	40.00
average	32.29	31.14	19.00

6. Conclusion

We present SVM algorithm with kernel method which is efficient especially for identification problems because of less learning data comparing with conventional approaches. We get the data of one user using the kernel function which is effective for classification of non-linear system. The numerical simulation and the experiment verified the usefulness of the proposed algorithm. Voice identification of a special user based on more than 2 kernels for the same words is under research.

REFERENCES

[1] H. Drucker et al., “Support Vector Regression Machines,” in Neural Information Processing Systems. Cambridge, MA:MIT Press, vol. .9, 1997.

[2] Waleed H. Abdulla, Vojislav Kecman, and Nik Kasabov, "Speech-Background Classification by Using SVM Technique".

[3] S. Gunn, “Support Vector Machines for Classification and Regression” ISIS Technica 1 Report, U. of Southampton, 1998.

[4] Christopher J.C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition ” Data Mining and Knowledge Discovery, 2, 121-167(1998).

[5] Kyung-Rae Cho, Jul-Ki Seok and Dong-Choon Lee, "Mechanical Parameter Identification of Servo Systems using Robust Support Vector Regression, " 2004. Proc. Advanced Motion Control, pp.588~593, 1996s.

[6] Yang Liu, Peng Ding and Bo Xu. “Using Nonstandard SVM for combination of speaker verification and verbal information verification in speaker authentication system.” In Proc. ICASSP 2002.

[7] Peng Ding; Zhenbiao Chen; Yang Liu; Bo Xu, “ASYMMETRICAL SUPPORT VECTOR MACHINES AND APPLICATIONS IN SPEECH PROCESSING ” ICASSP 2002. Volume: 1 , 13-17 May 2002 Page(s): I-73 -I-76 vol. 1.

[8] Dong Enqing; Liu Guizhong; Zhou Yatong; Zhang Xiaodi; “Applying Support Vector Machines to Voice Activity Dection” Signal Processing, 2002 6th International Conference on , Volume: 2 , 26-30 Aug. 2002.

[9] Dai Hee Lee, Suk Gyu Lee and Ju Hyun Park, “Design of a service robot with dual manipulators and stereo vision, CICS’ 03 page743~746.