IMPROVEMENTS ON COMMON VECTOR APPROACH FOR MULTI CLASS PROBLEMS

Rifat Edizkan[1], M. Bilginer Gülmezoğlu[1], Semih Ergin[1] and Atalay Barkana[2]

[1]Department of Electrical and Electronics Engineering, Osmangazi University
Batı-Meşelik, 26480, Eskişehir, Turkey
phone: + (90) 222-2393750, fax: + (90) 222-2200535, email: {redizkan, bgulmez, sergin}@ogu.edu.tr
[2]Department of Electrical and Electronics Engineering, Anadolu University
Muttalip, Eskişehir, Turkey
phone: + (90) 222-3213550, email: atalaybarkan@anadolu.edu.tr

ABSTRACT

In multi-class problems, within- and between-class scatters should be considered in classification criterion. The common vector approach (CVA) uses the discriminative information obtained from within-class scatter of any class. It has been shown that this classical CVA method gives high recognition rates in multi-class problems. In this study, improvements on the CVA method that consider both within- and between-class scatters are proposed and they are compared with the classical CVA. method. Although both methods give almost the same recognition rates on TI-digit database, they give better dimensionality reduction than the classical CVA method. The improved CVA methods also reduce both the processing time and the memory requirement of the classification parameters.

1. INTRODUCTION

The Common Vector Approach (CVA) has been introduced especially for speaker independent speech recognition not long ago [1-3]. The CVA is a subspace method based on calculation of the common vector for each class and the use of this vector in the recognition of classes. The common vector is a unique vector which represents the common properties of each class. CVA gave satisfactory results for the insufficient data case (n: vector dimension $\geq$ m: number of vectors in each class) in the isolated-word recognition, speaker recognition and fault detection of motors [2,4,5]. CVA was also applied to the isolated word recognition for the sufficient case ($n<m$) and again satisfactory results were obtained [6]. In the previous studies using the CVA, the difference and indifference subspaces are constructed by considering within-class scatters and, both within- and between-class scatters [7]. The projection of the average vector of any class onto its indifference subspace gives the common vector of that class which is used in the recognition problems.

To improve the CVA method for the solution of multi-class recognition problems, two different optimization criteria that consider both the within-class scatters and between-class scatters are proposed in this study. In the first optimization criterion, the distances of inter-class distribution are maximized in the indifference subspace obtained from the within-class scatters. In the second optimization criterion, the distances of intra-class distribution are minimized in the difference subspace obtained from the between-class scatters.

This study is a part of the project (200315053) supported by the Research Fund of Osmangazi University.

The results indicate that the improved CVA method and classical CVA method which considers only within-class scatters give the same recognition rates but the first one is superior to the second one in view of subspace dimensions. This dimensionality reduction method is superior to the well known Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) since these two other methods will reduce the dimensionality usually with some loss in recognition rates [8-10].

2. THEORY

In this paper, two optimization criterion derived from the CVA method are proposed.

2.1. Within Class–Between Class Optimization Criterion

In the first proposed criterion, the distances between the classes are maximized in the indifference subspace obtained from the within-class scatters.

Let us assume that m represents the number of feature vectors of each class, n represents the number of elements in each feature vector and $\boldsymbol{a}_i^c$ (i=1, 2,.., m) represents a feature vector in class c. Initially, an indifference subspace is obtained by considering within-class scatters. In this subspace within-class scatter of the class c will be close to the average of that class. The within-class optimization criterion is defined as:

$$\boldsymbol{F}_w^c = \sum_{i=1}^{m} \left\| \boldsymbol{P}_w^{c\perp} \left(\boldsymbol{a}_i^c - \boldsymbol{a}_{ave}^c \right) \right\|^2 . \qquad (1)$$

Let the eigenvalues of the within-class scatter matrix $\boldsymbol{\Phi}_w^c$ sorted in descending order $(\lambda_1 > \lambda_2 > \cdots > \lambda_k > \cdots > \lambda_n)$ and the eigenvectors denoted as $\boldsymbol{u}_j^c$. The eigenvectors associated with the smallest (n-k+1) eigenvalues span the indifference subspace [6]. The projection matrix for the indifference subspace can be given as

$$P_w^{c\perp} = \sum_{j=k}^{n} u_j^c u_j^{cT} . \quad (2)$$

For a better discrimination in this indifference subspace, the distances between the classes must be maximized. For this purpose, an optimization criterion F_1^c is proposed and defined as:

$$F_1^c = \sum_{i=1}^{m} \left\| P_{B_n}^c \left(P_w^{c\perp} a_i^c - P_w^{c\perp} a_{ave}^c \right) \right\|^2 . \quad (3)$$

The between-class scatter matrix $\Phi_{B_n}^c$ in the new difference subspace is defined as

$$\begin{aligned} \Phi_{B_n}^c &= P_w^c \left[\sum_{i=1}^{m} (a_i^c - a_{r,ave}^c)(a_i^c - a_{r,ave}^c)^T \right] P_w^{cT} \\ &= P_w^c \Phi_B^c P_w^{cT} \end{aligned} \quad (4)$$

where $a_{r,ave}^c$ represents the average of feature vectors of the rest of classes. The criterion F_1^c is maximized by taking the maximum eigenvalues of $\Phi_{B_n}^c$. Assume that the eigenvalues are in descending order and the new indifference subspace is spanned by the eigenvectors corresponding to the largest r eigenvalues of $\Phi_{B_n}^c$. The projection matrix $P_{B_n}^c$ for new difference subspace can be computed as

$$P_{B_n}^c = \sum_{j=1}^{r} y_j^c y_j^{cT} \quad (5)$$

where y_j^c represents the eigenvectors that span new difference subspace. Maximization of the criterion F_1^c is realized for each class separately. With this approach, the distances between the data of any class and the average of the rest of the classes are maximized in the indifference subspace of within-class scatter.

2.2 Between Class –Within Class Optimization Criterion

In the second proposed criterion, the compactness of within-class scatter is provided in the difference subspace defined by the between-class scatter. In this optimization criterion, first a difference subspace is obtained from the between-class scatter. The criterion in Eq.(6) gives a subspace in which the distances between the average of any class c and the average of the rest of classes are maximized.

$$F_B^c = \sum_{i=1}^{m} \left\| P_B^c \left(a_i^c - a_{r,\,ave}^c \right) \right\|^2 \quad (6)$$

In Eq.(6) P_B^c represents the projection matrix for the difference subspace. Assume that the eigenvalues of the between-class scatter Φ_B^c are in descending order. In this case, the difference subspace is spanned by the eigenvectors corresponding to the largest r eigenvalues [7].

For a better discrimination in this difference subspace, the distances between the data of any class and their averages must be minimized. For this purpose, an optimization criterion F_2^c is proposed and defined as:

$$F_2^c = \sum_{i=1}^{m} \left\| P_{w_n}^{c\perp} \left(P_B^c a_i^c - P_B^c a_{r,\,ave}^c \right) \right\|^2 . \quad (7)$$

Minimization of this criterion is obtained in the new indifference subspace spanned by the eigenvectors corresponding to minimum eigenvalues of the within-class scatter matrix $\Phi_{w_n}^c$ given as

$$\begin{aligned} \Phi_{w_n}^c &= P_B^c \left[\sum_{i=1}^{m} (a_i^c - a_{ave}^c)(a_i^c - a_{ave}^c)^T \right] P_B^{cT} \\ &= P_B^c \Phi_w^c P_B^{cT} . \end{aligned} \quad (8)$$

Assume that the eigenvalues are in descending order and the new indifference subspace is spanned by the eigenvectors corresponding to the smallest (n-k+1) eigenvalues of $\Phi_{w_n}^c$. The projection matrix for the indifference subspace is computed from

$$P_{w_n}^{c\perp} = \sum_{j=k}^{n} z_i^c z_i^{cT} \quad (9)$$

where z_i^c represents the eigenvectors that span the new indifference subspace. The criterion F_2^c is minimized for each class separately. With this approach, the distance between the data of any class and its average is minimized in the difference subspace of between-class scatters.

2.3 Decision Rules

The decision rules given in [11] are modified in this study. Therefore the following decision rule is used for the optimization criterion F_1^c given in Eq.(3),

$$K^* = \underset{1 \le c \le C}{arg\,min} \left\| P_{B_n}^c P_w^{c\perp} (a_x - a_{ave}^c) \right\|^2 . \quad (10)$$

If the distance is minimum for any class c, unknown vector a_x is assigned to class c. In the optimization criterion, while $P_w^{c\perp}$ makes the features of any class close to their average values, $P_{B_n}^c$ tries to maximize the inter-class distances.

For the optimization criterion F_2^c given in Eq.(7), the following decision rule is used in classification:

$$K^* = \underset{1 \le c \le C}{argmin} \left\| P_{w_n}^{c\perp} P_B^c (a_x - a_{r,\,ave}^c) \right\|^2 . \quad (11)$$

If the distance is minimum for any class c, unknown vector a_x is assigned to class c. In F_2^c, while the inter-class distances are maximized, the features of any class get close to their average values.

3. EXPERIMENTAL WORK

In the experimental work, the TI-digit database is used. At first, silence regions at the beginning and at the end of each repetition are removed by using energy and zero-crossing thresholds. After the each repetition is pre-emphasized, it is divided into 8 frames. The Hamming window is applied to each frame. The overlap between the frames is set to ¼ of the number of samples in each frame. Thirty-three root-melcep parameters are calculated for each frame. Then these parameters are stacked in order to construct the feature vector with the size of 330 for each repetition of each digit. This is called Variable Frame Length (VFL) method [12].

The experimental work can be divided into two parts:

i) In the criterion F_1^c, two projection operations are applied to test vector a_x. After the features of a class are projected onto the indifference subspace defined by within-class scatter, they are reprojected onto the next subspace obtained from the between-class scatter. In this case, the dimension of the final subspace is determined by the projection matrix of the between-class scatter. For the criterion F_1^c, the recognition rates of the training and test sets are given in Table 1 and Table 2 respectively. As seen from Table 1, the criterion F_1^c gives the recognition rate of 100% when one eigenvector corresponding to the maximum eigenvalue is used. However, the same recognition rate is obtained when the eigenvectors corresponding to minimum three eigenvalues of the within-class scatter matrix are used [13].

Table 1. Average recognition rates obtained using the criterion F_1^c for the training set.

Number of λ (Between-class)	Number of λ (Within-class)		
	1	5	10
1	18.18	99.97	100.00
5		73.52	100.00
10			90.88

Table 2. Average recognition rates obtained using the criterion F_1^c for the test set.

Number of λ (Between-class)	Number of λ (Within-class)				
	280	300	311	313	315
10	92.27	94.73	95.55	94.45	93.82
30	94.55	95.82	96.36	96.64	97.27
50	95.09	96.00	98.64	97.45	97.91
73	95.36	96.64	98.45	98.45	**98.73**
80	94.91	97.09	97.73	98.55	98.45
100	94.18	96.00	97.27	98.27	98.18

The maximum recognition rate of 98.73% is obtained for F_1^c in the test set when the eigenvectors corresponding to the largest 73 eigenvalues are used.

The maximum average recognition rate obtained using F_1^c is only 0.09 points less than the result obtained from only the within-class scatter. Therefore it can be assumed that they are approximately equal. In this case, the advantage of the criterion F_1^c is that the classification can be performed in lower-dimensional subspaces. While the best dimension of the subspace is 309 [13] for within-class scatter, the dimension of the subspace is now 73 for the criterion F_1^c. Classifications in a lower-dimensional subspace will decrease processing time and requires less memory for the classifier parameters.

ii) In the criterion F_2^c, the projections are obtained from between- and within-class scatters sequentially. In this case, the final dimension of the subspace is determined by the projection matrix of the within-class scatter. In Table 3, the recognition rates for the training set are obtained using the within-class scatter that are computed from the subspaces spanned by the eigenvectors of the between-class scatters.

Table 3. Average recognition rates obtained using the criterion F_2^c for the training set.

Number of λ (Within-class)	Number of λ (Between-class)				
	260	270	280	290	300
10	81.27	92.75	93.41	93.62	98.90
20	95.80	98.21	98.90	99.47	99.84
30	98.24	99.55	99.61	99.82	**100.00**
40	99.42	99.74	99.87	99.89	99.95
50	99.79	99.97	99.92	99.97	100.00
60	99.95	99.95	99.97	99.97	100.00
70	99.97	99.97	100.00	99.97	100.00

In Table 4, the recognition rates for test set are obtained using within-class scatter that are computed from the subspaces spanned by the eigenvectors of the between-class scatters.

Table 4. Average recognition rates obtained using the criterion F_2^c for the test set.

Number of λ (Within-class)	Number of λ (Between-class)				
	270	280	290	300	310
250	96.64	97.09	97.18	96.27	95.36
260	93.73	96.55	97.09	97.27	96.18
263	93.91	95.91	97.09	**97.73**	96.55
270	94.82	93.73	96.55	97.18	97.18
280		94.82	93.73	96.55	97.09

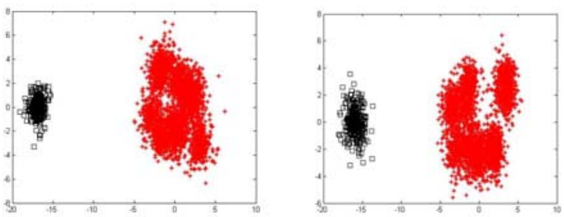
Just for comparison purposes the recognition rates are 99.6 and 98.18 with the HMM method and, 96.54 and 92.27 with the Heteroscedastic LDA (HDA) method for the training and test sets respectively

4. CONCLUSION

In the speech recognition, the desired performance cannot be obtained with the subspace methods when one uses only the within-class or between-class scatters. This situation is encountered especially when the number of classes increases. Therefore between-class scatters can be considered together with within-class scatters.

The best results were obtained for improved CVA that uses the optimization criterion $\boldsymbol{F}_1^c$ in which within- and between-class scatters are used. For this criterion, the recognition rate of 100% is obtained for the training set when the eigenvector corresponding to the maximum eigenvalue in the new difference subspace is used. For the test set, the recognition rate of 98.73% is obtained when the eigenvectors corresponding to maximum 73 eigenvalues in the new difference subspace are used. Although the recognition result of the improved CVA method is almost same as the results obtained from the classical CVA method, the improved CVA method provides dimensionality reduction much better than the classical CVA. For the improved CVA, the dimensions of the indifference subspaces are 1 and 73 for the training and test sets respectively. But the dimensions of the indifference subspaces in the classical CVA are 3 and 309 for the training and test sets respectively[13].

Figure 1(a-b) shows the scatters of any class and the rest of classes in the subspace spanned by the eigenvectors corresponding to largest 2 eigenvalues obtained from $\boldsymbol{F}_1^c$. The Figure 1 indicates how the classes are separated in two dimensional subspace. Note that the recognition rate in this subspace is obtained as 100% for the training set.



(a) (b)

Figure 1. The scatter of the classes in a 2-dimensional subspace obtained from $\boldsymbol{F}_1^c$. (a) “one” (□) and other classes (*), (b) “five” (□) and other classes (*)

The dimensionality reduction realized by the improved CVA $\boldsymbol{F}_1^c$ provides some advantages in designing isolated word recognizer which uses subspace techniques. Classification in a lower-dimensional subspace reduces the processing time and uses less memory space. The results indicate that the improved CVA can be easily used as recognition algorithms in real-time applications.

Although the other CVA method that uses $\boldsymbol{F}_2^c$ criterion gives similar recognition rates, it does not give comparable dimensionality reduction.

It would not be fair to compare HDA and HMM methods with the CVA for dimensionality reduction purposes since they are based on different mathematical derivations.

In future work, we continue our studies to improve the criterion $\boldsymbol{F}_2^c$. The criteria given in the paper will also be applied on databases that contain more classes.

REFERENCES

[1] M. B. Gülmezoğlu, V. Dzhafarov, M. Keskin and A. Barkana, “A novel approach to isolated word recognition”, *IEEE Trans. Speech and Audio Processing,* vol. 7, pp. 620-628, 1999.

[2] M. B. Gülmezoğlu, V. Dzhafarov and A. Barkana, “The common vector approach and its relation to the principal component analysis”, *IEEE Trans. Speech and Audio Processing*, vol. 9, pp. 655-662, 2001.

[3] A. Barkana, M. B. Gülmezoğlu, R. Edizkan, Ü. Künkçü and Ş. E. Künkçü, TÜBİTAK Project Report (Project No: EEEAG-82), 1995.

[4] M. B. Gülmezoğlu and A. Barkana, “Fault Detection of Motors Using Acoustical Measurements”, in *Proc. 6th Conference on Signal Processing and Applications*, Kızılcahamam, Turkey, 1998, pp. 290-295.

[5] M. B. Gülmezoğlu and A. Barkana, “Text-Dependent Speaker Recognition by Using Gram-Schmidt Orthogonalization Method”, in *Proc. IASTED Int. Conf. Signal Processing and Applications,* Canary Islands, 1998, pp. 438-440.

[6] M. B. Gülmezoğlu, V. Dzhafarov and A. Barkana, “Comparison of Common Vector Approach and Other Subspaces Methods in Case of Sufficient Data”, in *Proc. 8th Conference on Signal Processing and Applications*, Belek, Turkey, 2000, pp. 13-18.

[7] M. B. Gülmezoğlu, V. Dzhafarov, R. Edizkan and A. Barkana, ,“Common Vector Approach in Multi-Class Problems and Its Relation to Fisher’s Linear Discriminant Analysis”, in *Proc. 2nd Int Conf. on Electrical and Electronics Engineering (ELECO’2001)*, Bursa, Turkey, 2001, pp. 117-121,.

[8] C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.

[9] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, 1990.

[10] D.A Landgrebe, “Hyperspectral image data analysis”, IEEE Signal Processing Magazine, vol. 19, pp.17-28, 2002.

[11] R. Edizkan, M. B. Gülmezoğlu and A. Barkana, “A New Optimization Criterion in the Application of Common Vector Approach to Multi-Class Problem”, in *Proc. 11th Conference on Signal Processing and Applications,* İstanbul, Turkey, 2003, pp. 571-573.

[12] S. Ergin , M. B.Gülmezoğlu and A. Barkana, “Use Of Improved Feature Vectors In Affine Transformation For Robust Speech Recognition”, in *Proc. Int. Conf. on Information,* Çeşme, İzmir, 2004

[13] M. B. Gülmezoğlu, V. Dzhafarov, R. Edizkan and A. Barkana, “The Common Vector Approach and Its Comparison with Other Subspace Methods in Case of Sufficient Data”, *Submitted to Computer Speech and Language*, 2004.