DOI:10.26524/jms.2011.2

Feature matching by skpca with unsupervised algorithm and maximum probability in speech recognition

M.Pavithra, G.Chinnasamy, Azha.Periasamy, and S.Muruganand
Department of Electronics and Instrumentation, Bharathiar University.

Abstract

A Speech recognition system requires a combination of various techniques and algorithms, each of which performs a specific task for achieving the main goal of the system. Speech recognition performance can be enhanced by selecting the proper acoustic model. In this work, the feature extraction and matching is done by SKPCA with Unsupervised learning algorithm and maximum probability. SKPCA reduces the data maximization of the model. It represents a sparse solution for KPCA, because the original data can be reduced considering the weights, i.e., the weights show the vectors which most influence the maximization. Unsupervised learning algorithm is implemented to find the suitable representation of the labels and maximum probability is used to maximize the normalized acoustic likelihood of the most likely state sequences of training data. The experimental results show the efficiency of SKPCA technique with the proposed approach and maximum probability produce the great performance in the speech recognition system.

Keywords: Microfinance, women’s empowerment, Non Governmental Organization, Self Help groups.

I. Introduction

This paper deals with the applicability feature extraction and matching in speech recognition system using SKPCA (Sparse Kernel Principle Component Analysis) with unsupervised learning algorithm and maximum probability. Feature extraction of speech is the main step to match the speech data.The most commonly used feature extraction techniques used in speech recognition are Mel frequency cepstral coefficients (MFCC) [1], linear prediction coefficients-cepstral (LPC-cepstral) coefficients [1] and perceptual linear prediction (PLP) coefficients [2]. They have already been very well analyzed and their efficiency widely proved. The introduction of kernel based approach is to manipulate data into feature space. The main idea is to express the speech data in a higher dimensional space to generate possible discriminative features.

Kernel Discriminative Analysis (KDA), Kernel Principle Component Analysis (KPCA) [5] and Sparse Kernel Principle Component Analysis (SKPCA) are the some of the examples of kernel based approach. This kernel based approach is first applied by Supportive Vector Machines (SVMs) [3]-[4]. KPCA is non-linear method of approach to PCA.

The SKPCA is developed to solve the huge computational burden by generating the reduced data set through a likelihood maximization criterion.

Unsupervised Learning algorithm is provided the data points and no labels; the task is to find a suitable representation of the underlying distributing of the data. One major approach to unsupervised data is clustering. The unsupervised part is usually applied first to the data in order to make some assumptions about the distribution of the data, and then these assumptions are reinforced using a supervised approach [11], [12].

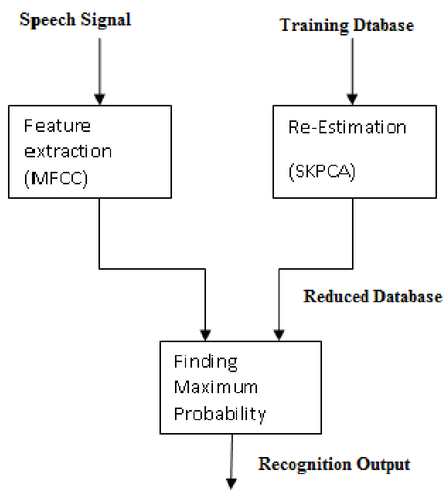

Figure 1. Block diagram of proposed study

As it is shown in Figure 1, the SKPCA technique can be separated in two blocks, the re-estimation and the KPCA block. The covariance matrix used in SKPCA approach is modelled as the weighted outer product of the training speech feature vectors plus an isotropic noise component, and these weights are up-dated by the re-estimation block. These weights generate the sparse solution for the KPCA, because they represent a measure of how well a specific training vector contribute to the likelihood maximization. Once obtained the reduced data, the common KPCA technique is applied and the representation of a feature test vector w is given by W_{skpca}. The Figure 1 also shows where the SKPCA is located in the speech recognition system as a whole.

Although the SKPCA generates a reduced training data, it requires the full original training data to evaluate the maximization step, which could be computationally unfeasible, depending on the training data amount. In order to solve it, an approach is proposed, where the original training data is clustered and the SKPCA is applied to these clusters. Despite this approach does not guarantee that the overall data maximum is reached, it will be shown by experimental results that SKPCA could surpass the performance of KPCA and standard feature extraction techniques. Maximum probability is introduced after getting the clustered training data. Most relative clusters of the original training data and test samples are matched accordingly.

The paper is structured as follows. In Section 2, a detailed evaluation of KPCA techniques are described, emphasizing the main points to reach SKPCA. The SKPCA is explained in Section 2.2, which comprises the weights re-estimation and the feature space representation. In Section 3, the objective approach is described and in Section 4, experiments are presented assuring the efficiency of SKPCA. Finally, Section 5 presents the conclusions of this work and ideas for future work are proposed, as well.

II. SKPCA Feature extraction

KPCA: Kernel functions are widely used to solve the problem of nonlinear mapping (Á) to a higher dimensional space, without using explicit mapping, which would be computationally unfeasible. If the function $k(x_i, x_j) = Á(x_i) \cdot Á(x_j)$ is a symmetric positive function that obeys the Mercer's condition [4], then it can be shown that this function represents the dot product of the variables xi and xj in the feature space. The kernel matrix K is defined as the matrix whose indexes are $(K)_{ij} = k(x_i, x_j)$.

The Kernel PCA is the technique which applies the kernel function to the PCA technique, in order to obtain the representation of PCA in a higher dimensional space [6]. In practice, the individual mapping of the data does not occur; the mapping is performed implicitly based on the dot products of the data, however this way of presenting the KPCA technique makes the procedure easier to be understood. Although the KPCA is a powerful technique, it has the disadvantage of requiring the full training data to calculate the kernel matrix [5]. The number of frames (full

training data) is used to reduce N frames, which are randomly picked up from the training data, as it was cited in [7]. Although this approach has shown an efficient performance [8], it does not use the overall information provided by the training data, i.e., when frames are randomly picked up, the training data is reduced without necessarily keeping any statistics of the original data set.
An approach where the frame reduction obeys the output probability maximization criterion was developed, and it is called SKPCA. It consists in estimating the feature space sample covariance for a noise component and the sum of the weighted outer products of the original feature vectors, which generate a sparse solution to KPCA.

SKPCA: In order to provide a solution for the previous mentioned disadvantage of the KPCA technique, an approach where the frame reduction obeys the output probability maximization criterion was developed, and it is called SKPCA. It consists in estimating the feature space sample covariance for a noise component and the sum of the weighted outer products of the original feature vectors, which generate a sparse solution to KPCA. This is obtained by maximizing the likelihood of the feature vectors.

$$CF = \sigma^2 I + \sum_{i=1}^{N} \omega_i \Phi_i^T$$
$$= \sigma^2 I + \Phi W \Phi^T \cdots\cdots\cdots\cdots (1)$$

Where W is a diagonal matrix composed by the adjustable weights $\omega_1, \ldots, \omega_M$, and σ^2 is an isotropic noise component, N $(0, \sigma^2 I)$, common to all dimensions of feature space. It was observed that fixing σ^2, and maximizing the likelihood under the weighting factors ω_i, the estimates of several ω_i are zero, thus realizing a reduced (sparse) representation of the covariance matrix. This approach was based on the probabilistic PCA (PPCA) formulation [9]. Further simplification of this formula is made by the re-estimation of the feature weights and SKPCA of the diagonal matrix can be obtained [10].

$$W_{skpca} = V^T{}_F \Phi\,(w) = \Lambda_K^{-1/2}\, \hat{U}^T{}_K\, k^T{}_W \ldots\ldots\ (2)$$

where $\hat{U}_K$ and Λ_K are defined, respectively, as the eigenvectors and eigen values of $W^{1/2}KW^{1/2}$and k_w represents the vector calculated by $k(w, x_i)$, where x_i corresponds to the non-zero weighted vectors represented in X.

III. Feature Matching

Objective approach of the SKPCA with Unsupervised algorithm: The proposed approach consists in making the SKPCA technique computationally feasible for a data set with a great number of samples, which is a usual situation in speech recognition. Generally, in speech recognition the amount of training data tends to be large, for example, in this work the training data comprises about 256 frames, and with this number of frames the SKPCA re-estimation is computationally unfeasible, once it depends on the kernel matrix K to calculate ∑.

Unsupervised Algorithm for training database: K-means clustering is proposed to divide the full training data into unsupervised clusters algorithm of L frames, and then merge the clusters forming new clusters of 2L frames, which are reduced to L frames by using SKPCA. The process is repeated successively until obtaining just one cluster of L frames, which is the final number of frames desired to represent the full training data. This L frames consists of their own cancroids.

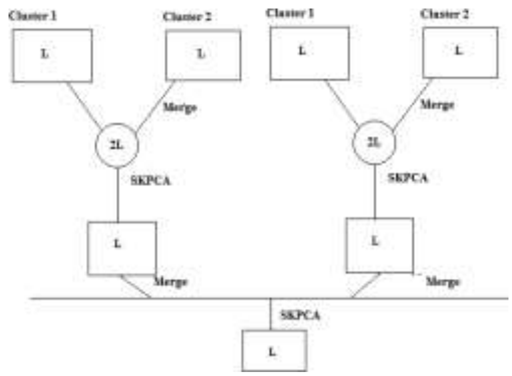

Figure 2. Block diagram of SKPCA study

The total number of steps necessary to reduce the l clusters of L frames to just one cluster, is given by step = log2l, where step is the number of steps. Considering this, the proposed approach does not guarantee to reach the over all data maximization, just individual cluster maximization.

Maximum Probability: Most relative clusters of the original training data and test samples are matched using their own individual covariance matrix. According to [13], the general form is

$$\mu` = A` \mu b` \quad ....(3)$$

The sample covariance matrix is given by,

$$C = 1/M \sum_{j=1}^{M} x_j x^T_j = M^{-1} X X^T \quad ....(4)$$

Where $X = [x_1 . . . x_M]$ represents the matrix of data. The M is number of centred observation of the dataset. Both training data and test samples covariance matrix are computed and matched according to their clusters.

IV. Experimental work

The efficiency of this technique is evaluated using speaker-independent isolated word recognition experiment was conducted. The experiment consisted in using a larger database, 200 Tamil words (one of the Indian language) with 120 speakers (60 males and 60 females) extracted database. The training data was composed of 10,000 utterances and the remaining 14000 utterances were used in test data.

Settings: The speech signals were obtained with 8000Hz of sampling rate, at a size of 16 bits per sample and dual channel. Pre-processing is considered as the first step of speech signal processing, which involves the conversion of analog speech signal into a digital form. It is a very crucial step for enabling further processing. The immediate next step is framing. It is a process of segmenting the speech samples obtained from the analog to digital (A/D) conversion into small frames with time length in the range of (20 to 40) milliseconds. Then, 61 Mel cepstral coefficients were extracted from each frame by using 25.6 ms Hamming windows with 10 ms shifts.

SKPCA method and maximum probability: The reduced number of training frames N applied to SKPCA with proposed approach experiment with 256 frames. The baseline features were obtained by applying 61 Mel cepstral co-efficient. The proposed approach was applied using L=256. The reduced number of frames N was obtained according to L, i.e., the training data set obtained by using L=256 was used in N=256, were obtained by merging the final clusters achieved in (step − 1) and (step − 2) steps. The error reduction of the overall best performance is 6.29% (SKPCA) against the best performances of the others techniques. The maximum probability method is applied to match the cancroids of each training feature data to test data. If both data are matching the corresponding feature is identified.

V. Conclusions and Future work

In this paper, 200 Tamil words (one of the Indian Language) recognition task were presented. From this, the effectiveness of SKPCA technique with unsupervised algorithm is confirmed in the case of feature matching. The importance of SKPCA is to improve the recognition accuracy of the system. In future SKPCA technique can be improved by introducing various type of unsupervised algorithm.

References

[1] L. R. Rabiner and B. Juang, *Fundamentals on Speech Recognition*, Prentice Hall, New Jersey, 1996.
[2] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *Journal of the Acoustical Society of America*, , no. 87, pp. 1738–1752, 1990.
[3] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer- Verlag, 1995.
[4] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, 1998.
[5]V. Roth and V. Steinhage, "Nonlinear discriminant analysis using kernel functions," *Advances in Neural Information Processing Systems*, pp. 568–574, 1999.
[6]B. Sch¨olkopf, A. Smola, and K.R. M¨uller, "Kernel principal component analysis," Advances in KernelMethods, pp.327–352, 1998.
[7]A.J. Smola, O.L. Mangasarian, and B. Sch¨olkopf, "Sparse kernel feature analysis," tech. rep., University of Wisconsin, 1999.
[8] A. Lima, H. Zen, Y. Nankaku, C.Miyajima, K. Tokuda, and T. Kitamura, "On the use of KPCA for feature extraction in speech recognition," IEICE Trans. Inf. & Syst., to appear.
[9] M.E. Tipping and C.M. Bishop, "Probabilistic principal component analysis," Journal of the Royal Statistical Society, pp.611–622, 1999.
[10] M.E. Tipping, "Sparse kernel principal component analysis," Advances in Neural Information Processing Systems, 2000.
[11] S.J. Cox and J.S.Bridle, "Unsupervised Speaker Adaptation by Probabilistic spectrum fitting," in proceeding ICASSP, pp. 294-297, 1989
[12] Glen Fung, "A comprehensive algorithm of basic cluster algorithm," pp.3 2001
[13] V.V. Digalakis, D.Rtischev, and L G Neumeyer, "Speaker adaptation using constrained estimation of Gaussian mixtures," IEEE Transactions Speech and Audio Processing, 3:357-366, 1995.

Received June 2011
Accepted July 2011
Final acceptance in revised form September 2011