

Discriminative transformation for speech features based on genetic algorithm and HMM likelihoods

Behzad Zamani^{1a)}, Ahmad Akbari¹, Babak Nasersharif^{1,2}, Mehdi Mohammadi¹, and Azarakhsh Jalalvand¹

¹ Audio & Speech Processing Lab, Computer Engineering Department, Iran University of Science & Technology, Iran

² Computer Engineering Department, Faculty of Engineering, University of Guilan, Iran

a) bzamani@iust.ac.ir

Abstract: Hidden Markov Model (HMM) is a well-known classification approach which its parameters are conventionally learned using maximum likelihood (ML) criterion based on expectation maximization algorithm. Improving of parameter learning beyond ML has been performed based on the concept of discrimination among classes in contrast to maximizing likelihood of each individual class. In this paper, we propose a discriminative feature transformation method based on genetic algorithm, to increase Hidden Markov Model likelihoods in its training phase for a better class discrimination. The method is evaluated for phoneme recognition on clean and noisy TIMIT. Experimental results demonstrate that the proposed transformation method results in higher phone recognition rate than well-known feature transformations methods and conventional HMM learning algorithm based on ML criterion.

Keywords: minimum classification error, genetic algorithm, speech recognition

Classification: Science and engineering for electronics

References

- [1] X. He and L. Deng, "A new look at discriminative training of Hidden Markov Models," *Pattern Recognition Lett.*, vol. 28, pp. 1285–1294, 2007.
- [2] B. H. Juang, W. Chou, and C. H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 3, pp. 257–265, 1997.
- [3] S. Kapadia, V. Valtchev, and S. J. Young, "MMI training for continuous phoneme recognition on the TIMIT database," *IEEE Int. conf. Acoust. Speech, Signal Process.*, vol. 2, pp. 491–494, 1993.
- [4] B. Nasersharif and A. Akbari, "SNR-dependent compression of enhanced Mel sub-band energies for compensation of noise effects on MFCC fea-

- tures,” *Pattern Recognition Lett.*, vol. 28, no. 11, pp. 1320–1326, 2007.
- [5] M. Loog and R. P. W. Duin, “Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 732–739, 2004.
- [6] X.-B. Li, J.-Y. Li, and R.-H. Wang, “Dimensionality reduction using MCE-optimized LDA transformation,” *IEEE Int. conf. Acoust. Speech, Signal Process.*, vol. 1, pp. 137–140, 2004.
- [7] A. De La Torre, A. M. Peinado, A. J. Rubio, V. E. Sanchez, and J. E. Diaz, “An Application of Minimum Classification Error to Feature Space Transformation for Speech Recognition,” *Speech Commun.*, no. 20, pp. 273–290, 1996.
- [8] D. E. Goldberg, “*Genetic Algorithms in Search, Optimization and Machine Learning*,” Addison-Wesley, New York, 1989.

1 Introduction

Speech recognition systems include two main components: feature extraction and classification. A well-known and successful classification approach for speech recognition is Hidden Markov Model (HMM). The HMM parameters are usually learned via maximum likelihood (ML) criterion based on expectation maximization algorithm [1]. Improving of parameter learning beyond ML has been performed based on the concept of discrimination among classes in contrast to maximizing likelihood of each individual class. Thus, discriminative training methods such as minimum classification error (MCE) [1, 2] and maximum mutual information (MMI) [3] have been proposed.

In the feature extraction module, the useful discriminative information is extracted from speech signal such that the HMM classifier can recognize different speech units including phones, tri-phones, syllables or words. The most widely used and successful speech features are Mel-frequency cepstral coefficients (MFCC). However, MFCC are not optimal for speech features discrimination. Furthermore, their performance degrades in the presence of additive noise. Hence, several methods have been suggested for extracting more discriminative and robust MFCC [4] such as Linear Discriminant Analysis (LDA), Heteroscedastic LDA (HLDA) [5], and Principal Component Analysis (PCA). The drawback of these transformations is that their optimization criteria are different from the classifier’s minimum classification error criterion, which can potentially corrupt the classifier performance. There are several methods to overcome this drawback. In some approaches, feature extraction and classification are conducted jointly based on a consistent criterion [6]. In this paper, we improve HMM training and class discrimination using a feature transformation based on genetic algorithm (GA).

2 Feature Transformation Based on Genetic Algorithm for Increasing HMM Likelihoods

In some of feature transformation methods like PCA, LDA and HLDA, we multiply original features by a transformation matrix obtained through the

approaches as:

$$O = W \times X \quad (1)$$

where W , X and O are transformation matrix, original features and transformed features, respectively. W is obtained based on the transformation method criterion.

In this paper, we propose a method for obtaining transformation matrix W in order to increase HMM likelihoods in its training phase and so increase the discrimination between classes. We achieve this matrix using the genetic algorithm. We calculate this matrix as:

$$\arg \max_{W_{p,g}} \frac{1}{M} \sum_{i=1}^M \log P_{i,k}(W_{p,g} \times X_i | \lambda_i) \quad (2)$$

where, $W_{p,g}$ is the p -th transformation matrix (the p -th population member or the p -th individual) in generation g ; X_i contains all original observation vectors of the i -th class; λ_i denotes HMM of class i ; M is number of all classes and $P_{i,k}$ indicates likelihood for model i in k -th iteration of HMM training phase using Baum-Welch algorithm. Fig. 1 shows the proposed GA based method more clearly. The presented parts in the gray box in Fig. 1 show our GA method proposed for the improving HMM training phase and increasing likelihoods. You can see that If we remove the gray box in the figure and consider W as identity matrix, we will have the same conventional HMM training method.

Using the genetic algorithm, we should find the best transformation matrix (best population member) in the g -th generation that maximizes mentioned average. This means that HMM training phase should be repeated (*population size* \times *number of all generations*) times. We describe the implemented genetic algorithm in the following.

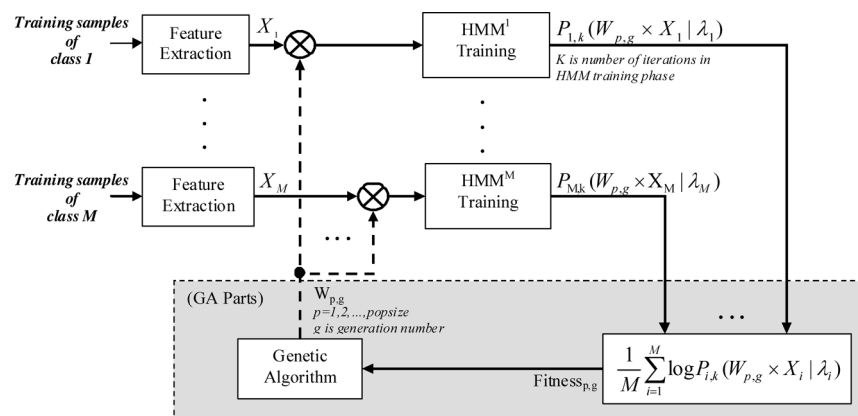


Fig. 1. The proposed method based on GA and HMM likelihoods

2.1 Representation

In the first step of defining a GA, we should create a relation between the real world and the GA world. In this way, a bridge between the original

problems context and the problem solving space is made. For this purpose, we should define population according to our problem space. Each member of population is named as an individual that should contain the parameters related to our problem. We define our individual (chromosome) structure as vector with length n , where n is the feature vector dimension. This chromosome includes the main diagonal of our transformation matrix supposed to be diagonal in this paper. Other elements of transformation matrix are assumed zero.

2.2 Population initialization

A part of first population is constructed by randomly generated individuals. Another part of population is build from MCE matrix for feature transformation [7], plus to PCA, LDA and HLDA matrices (heuristic initialization [8]). Using this approach, we guide GA to the regions in the search space that contain better results.

2.3 Fitness function

The role of the fitness function is to evaluate the quality of solutions. Based on these evaluations, better chromosomes are selected for the new population. Our proposed fitness function is the average log-likelihood of all HMMs in the k -th iteration. Our fitness function is maximized to improve the Hidden Markov Models training for each class. We obtain our fitness function in three steps as follows:

1. Each matrix in the current population is applied to the train samples.
2. HMM models are trained based on these transformed samples.
3. The log-likelihood values for all HMMs are obtained in the k -th iteration. Subsequently, fitness function is calculated based on equation (3) for the current matrix in the population.

$$fitness_{p,g} = \frac{1}{M} \sum_{i=1}^M \log P_{i,k}(W_{p,g} \times X_i | \lambda_i) \quad (3)$$

Where, λ_i is the i -th HMM model, X_i is the train samples of the i -th class, $P_{i,k}(X_i | \lambda_i)$ is the likelihood value of the i -th HMM model in the k -th iteration and M is the number of classes. The fitness calculation method can also be seen form Fig 1. Using this fitness function, GA finds the best matrices that maximize average log-likelihoods in order to improve HMM training and so class discrimination.

2.4 Recombination (Crossover)

Crossover operator is a variation operator which generates new individuals in the genetic iterations. Crossover often is a probabilistic process that exchanges information between some (usually two) parents in order to generate some new children. Here, we use “Uniform” crossover with a fixed probability named p_u for occurring crossover. In the uniform crossover, individual genes

are exchanged between two parents with a determined probability. Thus, new children inherit some genes from the first parent and some genes from the second parent, uniformly. The equation (4) indicates this crossover:

$$W_{p,g}^x = \begin{cases} W_{p,g}^x & p_x \geq 0.5 \\ W_{q,g}^x & p_x \leq 0.5 \end{cases} \quad (4)$$

Where, $W_{p,g}^x$ indicates the element x on the main diagonal of Matrix $W_{p,g}$ in the p -th population member of the g -th generation. p_x is a random number between 0 and 1 used as genes exchange probability. Using Eq. (4), we generate two offspring for each two parents. So, population size is not changed.

2.5 Mutation operator

Mutation operator is another variation operator that randomly changes one or more genes of the chromosome. The mutation operator prevents the genetic population from converging to local optima and introduces new possible individuals which do not exist in the previous population. The mutation is carried out according to the mutation probability. Three Mutation methods are used in this work as follows.

- *Random mutation:* This mutation assigns a new random value to a random selected gene of a chromosome. In the other words, this type of mutation introduces new individuals to the population.
- *Creep mutation:* In this method, by adding a small (positive or negative) value to a selected gene, a new value is assigned to the gene to generate small changes in the chromosome. Therefore, chromosomes change more slowly than random mutation case. Equation (5) shows this mutation where R is a given integer value, selected equal to 10 in this work.

$$W_{p,g}^x = W_{p,g}^x \pm r \quad r \in [0, R] \quad (5)$$

- *Swap mutation:* This operator simply exchanges the position of two random selected genes. Based on this reconfiguration, a new mutant child is created. This mutation is done as:

$$W_{p,g}^x \leftrightarrow W_{p,g}^y \quad (6)$$

Where $W_{p,g}^x$ indicates the element x on the main diagonal of Matrix $W_{p,g}$ in the p -th population member of the g -th generation and $x \leftarrow y$ denotes the replacement of x with y .

2.6 Selection operator

The selection process selects individuals from the population based on their fitness values [8]. In this paper, we use the binary tournament selection in which, two individuals are chosen from the population at random and the individual with better fitness is selected for the next population. Using the above framework, feature space is transformed by GA in order to increase HMM likelihoods in its training phase.

3 Results

We name our proposed transformation method as GAML in the experiments. GAML is using MCE, LDA, HLDA and PCA matrices as its initial population. Consequently, the GAML matrix contains the information about minimum classification error. So, it can be compared with these methods, however its fitness function is based on ML. The training of HMM begin from the start at each generation of GAML method. So, the comparison between GAML and other methods perform in fair conditions.

We report phoneme error recognition rate (PER) on TIMIT database. In our experiments, we use 39-dimension feature vectors consisting of energy, 12 MFCCs and their first and second order derivatives. The features were normalized to have mean zero and standard deviation one over TIMIT training set. We don't reduce features vector dimension in our methods. In addition, we use diagonal covariance HMMs with 3 states and 16 Gaussian mixtures per state. We use TIMIT train set for training HMMs and utilize its test set for recognition experiments. For showing performance of our methods in noisy conditions, we added noise to TIMIT test set. We selected three noises from NOISEX92 database: white, pink and factory1. Then, we added these noises to all of TIMIT test set sentences with different SNR values of 0, 5, 10, 15 and 20 dB. HMMs are trained using clean training set of TIMIT database.

Table I. Average phoneme error recognition rate (PER) on clean TIMIT test set and different SNR values for three added noises to TIMIT test set

	White	Pink	Factory1	Clean Test Set
MFCC	45.94	48.49	49.33	71.69
MCE	46.14	48.5	49.28	71.83
PCA	45.67	47.97	48.89	70.18
LDA	46.66	49.36	49.91	70.25
HLDA	46.95	49.43	49.91	70.39
GAML	47.17	49.51	50.06	71.84

As shown in Table I, GAML performs better than MFCC and so conventional HMM parameters learning based on ML criterion and Baum Welsh algorithm. This is due to its ability in discriminative transformation. Furthermore, GAML has higher recognition rate than PCA, LDA and HLDA methods that don't use from classifier (HMM) criterion. This happens due to its use from classifier (HMM) information in determining transformation matrix. In addition, GAML performs better in comparison to MCE that only transforms feature as in [7]. This can be due to the evolutionary property of GA that use MCE matrix as its initial population.

4 Conclusion

In this paper, we propose a feature transformation method to increase HMM likelihoods in its training phase for better class discrimination. The method

is based on genetic algorithm. In this approach, we computed a diagonal matrix W in order to change the input feature spaces to maximize the average of HMM likelihoods in its training phase. We haven't any dimension reduction in this case. One part of initial population of genetic algorithm includes MCE, PCA, LDA and HLDA transformation matrices, while other parts include matrices generated randomly. Results on TIMIT show that the proposed method improves the conventional HMM parameters learning method based on maximum likelihood measure. Additionally, it outperforms the transformations such as PCA, LDA and HLDA that don't notice to classifiers error. Furthermore, it has better performance than MCE.