Vol. 8(27), Jan. 2018, PP. 3753-3763

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

Unsupervised Feature Selection for Phoneme Sound Classification using Genetic Algorithm

Mohammad Mahdi Faraji[1*], Saeed Bagheri Shouraki[2], and Ensieh Iranmehr[3]

[1,2,3] *Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran*

***Corresponding Author's E-mail:** mmfaraji@ee.sharif.edu*

Abstract

This paper proposes a new method based on Genetic Algorithm for feature selection in phonemes sound classification. Biological studies have shown that human's ear is sensitive to different resonant frequencies because of ear's hair cells. Thus, we propose a technique in which genetic algorithm is used to extract audio features similar to human's ear in order to achieve better classification. In this paper, genetic algorithm is used in order to select appropriate individual's features in order to classify sound signals accurately. Each individual consists of genes indicating the resonant frequencies inspired from human cochlea hair cells. Then, feature extraction is done by using individual's information. Moreover, a fitness function by using classification method based on nearest neighbor is used in order to evaluate each individual of population. Furthermore, by using the proposed genetic algorithm, best individual's features can be found. In order to evaluate this proposed method, a database which consists of 500 samples for each 12 different phoneme classes is created in this paper. The proposed algorithm is compared with an existing typical audio feature selection based on MFCC and the proposed algorithm achieves much better classification accuracy in comparison with MFCC based feature selection method. During generations, the fitness value shows remarkable improvement of sound classification accuracy.

Keywords: Genetic Algorithm, Sound Classification, Unsupervised Feature Selection, MFCC, PSO.

1. Introduction

The classification of signals is a research topic that has attracted a lot of interests since its beginning [10, 5]. In general, signal classification is usually divided into three main stages: preprocessing, feature extraction and classification. This paper focuses on the second stage more than the others. Feature extraction is one of the most important stage, in which various features are extracted from a signal. Then, these features are used as inputs to a classification system. In the stage of classification, a classifier generates an output which determines class of the signal. In various audio classification problems, features are usually extracted based on Mel Frequency Cepstral Coefficients (MFCC), without applying optimization approach [9]. Several evolutionary based optimization techniques seem to be very promising for such tasks [6, 8, 2]. In this paper, we study the audio feature selection based on Genetic Algorithm (GA) which is an evolutionary technique. Evolutionary algorithm such as genetic algorithms (GAs) [3] has attracted much interest in various applications [16]. In a GA, each solution to a problem is called an individual. A set of solutions is called a population. Genetic algorithms are commonly used to generate high-quality solutions to optimize and search problems by relying on bio-inspired operators such as mutation, crossover and selection. These operators act similar to their

Article History:

DOI: IJMEC/10.225163

Received Date: Mar. 16, 2017

Accepted Date: Aug. 21, 2017

Available Online: Jan. 05, 2018

counterparts in the natural evolution: in the population, the top individuals have a higher chance to reproduce themselves. By using these operators, the population follows an iterative process which goes through different states, each of which is called a generation. As a result of this process, the population is expected to reach a generation that contains a good solution to the problem. One of the most popular pattern recognition application for GAs is feature selection. Although many authors support feature selection by GAs [18, 14], others are hesitant [13], and inform that the results are often not as good as expected, compared with other feature selection algorithms [1].

This paper is inspired of the results of biologist's studies about how human's ear acts. As they mentioned, cochleae which is the little spiraling tubes in our inner ears, separates the frequencies of incoming mechanical wave. Human cochlea is equipped with thousands of hair cells which vibrate with the surrounding liquid. Each hair cell has different resonant frequencies. Depending on the location in the cochlea that vibrates, different nerves fire in order to inform the brain that certain frequencies are present [17]. Moreover, the cochlea cannot realize the difference between two closely spaced frequencies. For this reason, [15] take clumps of periodogram bins and sum them up to get an idea of how much energy exists in various frequency regions. Thus, by using this concept and GA, we assume that each individual at initial is represented by a random filter-bank called individual's features in this paper.

This paper proposes a new technique for sound classification in which GA is used to automatically extract individual's features. These frequency domain features contains individual resonant frequencies and bandwidths, which act as filters. These filters are multiplied to input audio spectrum in order to achieve feature vector which is heard by human ear. By means of proposed algorithm for classification which is based on nearest neighbor, each solution is evaluated by its fitness function (FF). Among all the solutions, some of the top solutions are selected as the parents of next generation. Finally, the solution, which has a better fitness value, is considered as optimal solution [7]. In this way, the best individual's features are chosen in order to classify the sound signal.

The rest of this paper is organized as follows: Section 2 explains the process of finding optimal solution in order to achieve better classification. Section 3 discusses the feature selection technique using the genetic algorithm and its procedure. Experimental results of the proposed method are presented in Section 4. Finally, the conclusion is provided in Section 5.

2. The Structure of Model

The aim of this paper is to increase the performance of sound classification by using genetic algorithm in feature selection. This paper tries to find optimum solution of the population through generations. "Figure 1" shows overall process of finding best solution. At first, sound signal has to be filtered by individual's filter-bank in order to create feature vector of each individual. For this purpose, Fast Fourier Transform (FFT) of signal is calculated and then it is multiplied by individual's filter-bank.

By creating 3-fold cross validation, test and train data are separated in order to compute classification error. In training step, the centroid of each class is computed by averaging of all training data. A classifier based on nearest neighbor is applied in order to determine the class of each test data. After that, the actual and computed classes of test data have to be compared and the error of classification of each individual is counted. This procedure is done for every individuals of a population. In this way, fitness value of every individual of the current generation is calculated.

Then, by using selection method, 40 individuals which their fitness values are better than the others are chosen. By means of crossover and mutation operators, next generation is produced. Termination criteria are considered to follow iterative process or ending the process. When the criterion is satisfied, the individual which has the best fitness value is considered as best solution of this problem. Otherwise, the process on next generation is done. Next section describes the details of proposed algorithm based on GA.

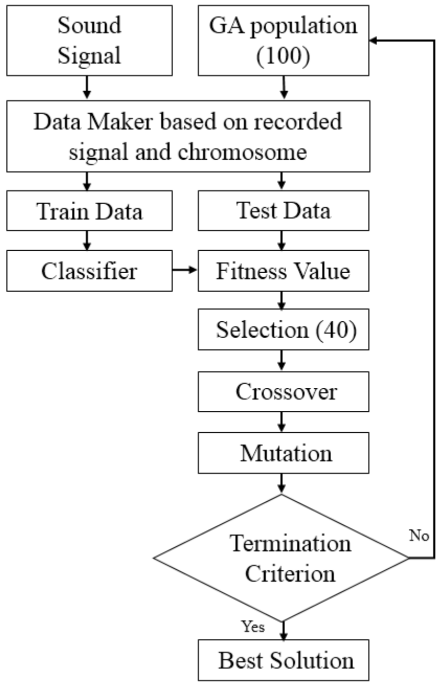


Figure 1: Overall process of finding best solution

3. Genetic Algorithm

By using GA, a population of individuals each of which has a set of genes, is evolved toward better solutions. In this paper, the evolution starts from a population of randomly generated individuals. In each generation, the fitness value of every individual in the current generation is evaluated. After choosing the more fit individuals from the current population, each individual's filter-bank of next generation is formed by modifying its parent's filter-banks. The new population is then used in the next iteration of the algorithm. The algorithm terminates when it reaches to a maximum number of generations.

In order to describe the proposed method in this paper, a genetic representation of individuals and a fitness function to evaluate individuals have to be introduced. When the genetic representation and the fitness function are defined, a GA initializes a population of individuals and improves it through repetitive mutation, crossover, and selection operators. In the following, the proposed genetic representation, fitness function, crossover, mutation and selection operators are introduced.

3.1. Genetic representation

A standard representation of each individual is as an array of bits [4]. In this paper, we consider the resonant frequencies of human's ear and their bandwidth as individual's filter-bank. Because of the range of human hearing frequency, these frequencies are assumed in the range of 20Hz to 15 kHz. As it can be seen from "Figure 2", the individuals are considered by using the frequencies and bandwidths. In "Figure 2", 8 features for each individual are shown which the area under each feature is the same. In order to represent each individual, we have to express feature with a vector which the first 10 bits indicates resonant frequency and second 6 bits indicates its bandwidth. Thus, individual' chromosome represents by an array with the size of $feature\_num \times 16$. The i^{th} row of individual's chromosome determines the index number of individual's feature [11].

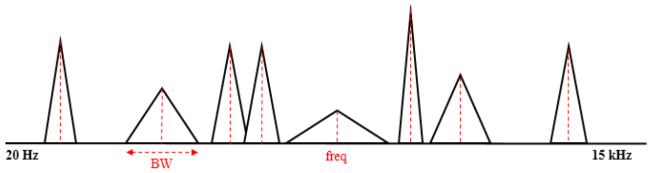


Figure 2: An example of individual's filter-bank

3.2. Fitness Function

The fitness function was used to indicate the goodness of the individuals. In this paper, in order to classify sound signal, we use an algorithm based on nearest neighbor for sound classification. As it can be seen from "Algorithm 1", the fitness value of each individual is calculated by using this classifier. The FF is a type of objective function used to evaluate each individual. In every generation, the fitness values of all individuals in population are calculated. "Algorithm 1" shows the pseudo code of calculating fitness value of each particle.

Algorithm 1: The pseudo code of Fitness Function

Inputs: *Individuals, Train and Test Data.*

Output: *Fitness Value of individual which is between 0 and 1.*
(The smaller value of fitness represents better fitness)

Computing fitness value:

- *Extract feature vector based on individual's features for both Train and Test Data.*
- *Use Train feature vectors and find average of them for each class of Train Data (centroids).*
- *Classify every Test Data based on the nearest centroid.*
- *Return F = Error / N_Test as a fitness value in which Error is the number of misclassified Test Data. And N_Test is the total number of Test Data.*

In '(1)', the average of training feature vectors of each class which is shown by $\overline{X_c}$ is computed where c is the index number of classes. In this equation, n_i is the number of training feature vectors and $X_{i,c}$ is i^{th} training feature vector of c^{th} class.

$$\overline{X_c} = \frac{1}{n_i}\sum_{i=1}^{n_i} X_{i,c} \tag{1}$$

In order to classify test data, we use '2'. In '(2)', the k^{th} test feature vector which is shown by X_k is compared to center of each class which is shown by $\overline{X_c}$. Since $\overline{X_c}$ and X_k are feature vector with size of j, we have to use Mean Absolute Difference (MAD) of these two vectors where n_j is the number of features of feature vector.

$$MAD(X_{k,c}) = \frac{1}{n_j}\sum_{j=1}^{n_j} |X_k(j) - \overline{X_c}(j)| \tag{2}$$

Then, the MAD of each test feature vector for every classes is computed. The minimum value of MADs for a test feature vector is found in order to classify k^{th} test data into the class which has minimum value. After all, the actual and the computed class of test data is compared to find total error and as can be seen from this algorithm, the fitness value is computed by using this error.

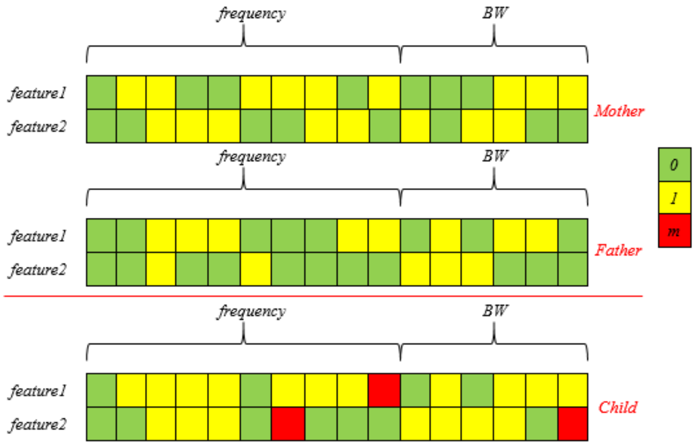
By calculating fitness value of each individual in population, genetic algorithm evaluates the individuals and by selection operator, finds some better individuals as parents of next generation. In the next subsections, how to generate next generation by using the operators is explained.

3.3. Crossover and mutation operators

Crossover is a genetic operator used to vary the individuals from one generation to the next. Many crossover techniques exist for organisms [4]. The patterns of the crossover operator are divided into one point, two points, and a uniform crossover.

In this paper, a crossover operator is used to produce a child's chromosome from its parent's chromosomes. The proposed crossover technique which is based on uniform crossover is described in this section. The uniform crossover uses a fixed mixing ratio between two parents. Unlike single- and two-point crossover, the uniform crossover enables the parent chromosomes to contribute the gene level rather than the segment level.

Because of different resonant frequencies of each individual's features, these frequencies and their bandwidth are translated into array of bits which is considered as chromosome. Assume mother and father's features are shown by X and X' respectively. First of all, by using genetic representation, the features of parents have to be translated into arrays of bits. In this way, two arrays of genes are constructed each of which determines chromosome. Children's chromosomes (offspring) are generated from their parent's chromosomes. As mentioned above, in this paper each chromosome indicates individual's features and it is an array with the size of (number of features×16) bits. In order to use uniform crossover, two arrays are contributed with a probability of 0.5. After crossover modifies parent's genes to form a new generation, the array, which is indicated offspring, has to be translated into child's features X''.



Fiqure 3: The representation of crossover and mutation operators

After a crossover is performed, mutation takes place in order to prevent the population of chromosomes from beginning too similar to each other. In other word, a Mutation technique is used to maintain genetic diversity from one generation of a population of genetic algorithm chromosomes to the next.

In this paper, Mutation changes randomly the new offspring. For binary encoding which is used in this study, we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1. "Figure 3" represents how to generate an offspring from its parent's chromosomes. In this figure, just two features are considered for indicating a chromosome due to its simplicity.

As it can be seen from "Figure 3", two chromosomes which are indicating by 32 genes are considered as parents and then crossover technique is applied to them in order to produce a child's chromosome. Next, mutation takes place in almost 10 percent of genes which is shown in "Figure 3" by red squares. By the means of mentioned operators, a new chromosome is generated.

An example of parents' filter-bank by considering 8 features has been shown in "Figure 4". In this figure, parents and a generated child's filter-bank are shown.

In this paper, Mutation changes randomly the new offspring. For binary encoding which is used in this study, we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1. "Figure 3" represents how to generate an offspring from its parent's chromosomes. In this figure, just two features are considered for indicating a chromosome due to its simplicity.

As it can be seen from "Figure 3", two chromosomes which are indicating by 32 genes are considered as parents and then crossover technique is applied to them in order to produce a child's chromosome. Next, mutation takes place in almost 10 percent of genes which is shown in "Figure 3" by red squares. By the means of mentioned operators, a new chromosome is generated.

An example of parents' filter-bank by considering 8 features has been shown in "Figure 4". In this figure, parents and a generated child's filter-bank are shown.

After using crossover and mutation operators, we have to select some of the superior solutions and build next generation from these 40 top individuals. Then, these 40 top individuals mate randomly.

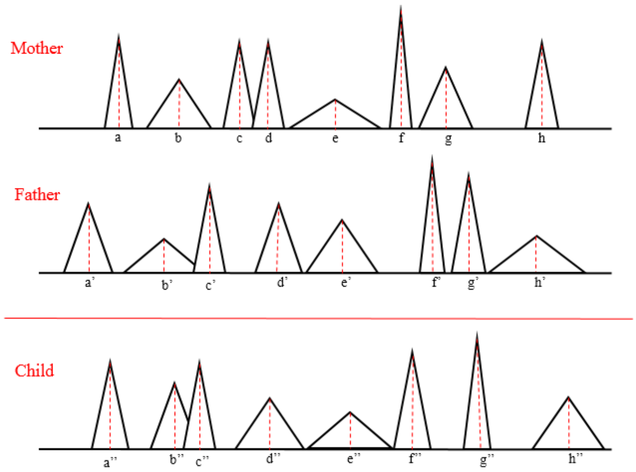


Figure 4: Generate child's filter-bank from its parents by using the proposed crossover and mutation operators

4. Experimental Results

In order to evaluate the proposed algorithm, a database which consists of 12 phonemes with a single speaker has been created. For each phoneme in the database, 500 tracks with the length of 4096 samples have been generated. Sampling rate is 96 Ksample/sec. Therefore, the length of each track is about 42 milliseconds. In other words, audio signal in each track can be assumed as a stationary signal because it can be presumed that human audio spectrum does not change over 42 milliseconds. These tracks are randomly divided into two sets: train and test data sets. "Figure 5" shows 4 different classes of phoneme tracks in frequency domain.

For evaluating the proposed algorithm, we consider different number of features which the length of each of them is 16 bits. Each bit is considered as a gene. Several experiments have been done with the aim of finding which parameters are the most appropriate, and "Table 1" summarizes these appropriate parameters which have been used in this paper.

Table 1: The parameters of used Genetic Algorithm

Parameter	Value
Number of generation	20
Population size	100 individuals
Chromosome Length	(number of feature$\times$16) genes
Crossover	Uniform
Mutation	10% detour from parent's genes
Selection	Random selection from 40 top individuals

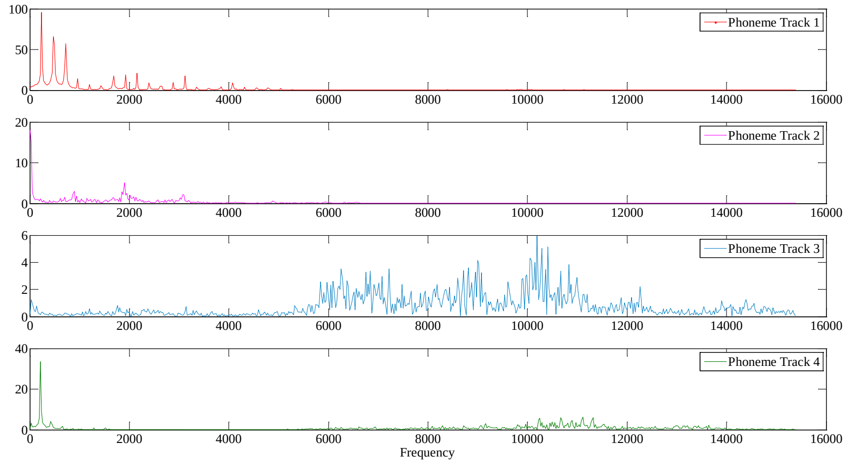


Figure 5: 4 different classes of phoneme tracks in frequency domain

In this paper, as mentioned in previous section, we use a classifier based on nearest neighbor in order to classify 12 phonemes and achieve fitness values by using the proposed algorithm. "Figure 6" shows the behavior of fitness curve with the mean and best value of the fitness function over each generation by considering 8 features. As shown in "Figure 6", the best fitness value is occurred in the 6^{th} generation with the value of 0.018.

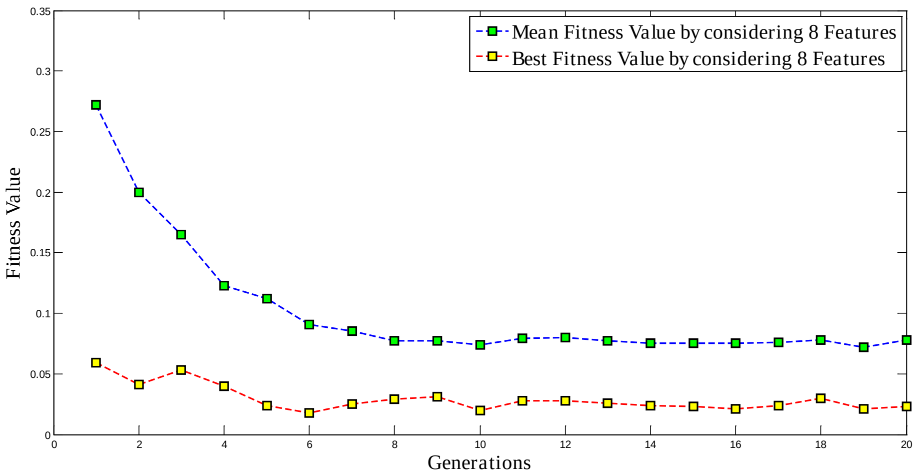


Figure 6: Fitness curve between the best and mean fitness value over each generation by considering 8 features for each individual

Moreover, in order to evaluate this proposed algorithm, we consider different number of features for each individual and then find the best fitness. "Figure 7" shows fitness curve between the best and mean fitness value by considering 16 and 32 features. As it can be seen from this figure, the best fitness value by considering 16 features is occurred in 10^{th} generation with the value of 0.006. And by considering 32 features, the best fitness value is occurred in 6^{th} generation with the value of 0.002. Furthermore, descending graph of mean fitness value shows that the average fitness value of all individuals is going to get better by producing a new generation.

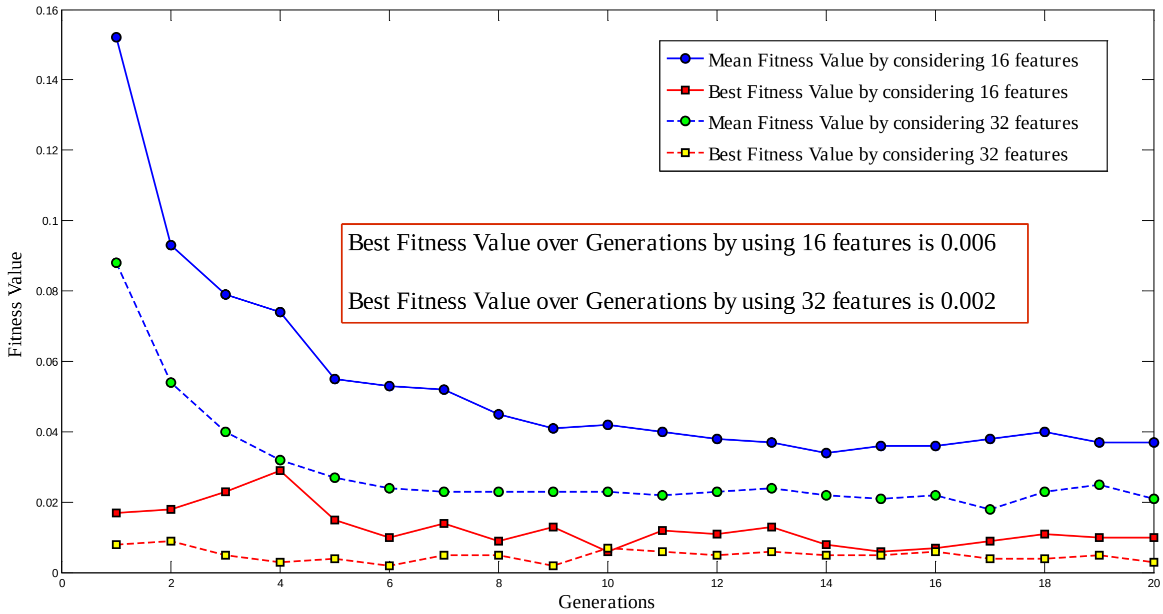


Figure 7: Fitness curve between the best and mean fitness value over each generation by considering 16 and 32 features for each individual

The best fitness value in each generation for different number of features is shown in "Table 2". Moreover, the best fitness value over generations for different number of features is shown in "Table 3". Furthermore, this proposed algorithm is compared with MFCC method [12], uniform feature selection and a similar algorithm based on Particle Swarm Optimization (PSO) [6].

Table 2: Best Fitness Value for Each Generation considering Different Number of Features

Number of Generation	Best Fitness Value		
	8 Features	*16 Features*	*32 Features*
1	0.059	0.017	0.008
2	0.041	0.018	0.009
3	0.053	0.023	0.005
4	0.040	0.029	0.003
5	0.024	0.015	0.004
6	0.018	0.010	0.002
7	0.025	0.014	0.005
8	0.029	0.009	0.005
9	0.031	0.013	0.002
10	0.020	0.006	0.007
11	0.028	0.012	0.006
12	0.028	0.011	0.005
13	0.026	0.013	0.006
14	0.024	0.008	0.005
15	0.023	0.006	0.005
16	0.021	0.007	0.006
17	0.024	0.009	0.004
18	0.030	0.011	0.004
19	0.021	0.010	0.005
20	0.023	0.010	0.003

In uniform feature selection, features are selected uniformly in the predefined frequency range. Despite using the same classifier for three different feature selection methods, but it can be seen from Table III, MFCC can select better features than uniform feature selection and our proposed algorithm selects much better features than MFCC. Moreover, by comparing the results with the results of [6], it is shown that the performance of these two algorithm is similar to each other.

As it can be seen from “Table 2” and “Table 3”, the accuracy of classification gets better by increasing number of features and the best fitness in the proposed algorithm is much better than the typical audio feature selection based on MFCC. In other words, the proposed genetic algorithm selects the best features which results in minimum classification error. This improvement is because of the fact that in the proposed method, feature selection is adapted with classifier properties over generations.

Table 3: Best Fitness Value for Different Number of Features

Evaluation parameters	Number of features			
	4	8	16	32
Fitness Value based on Uniform features	0.363	0.258	0.108	0.021
Fitness Value based on MFCC	0.177	0.048	0.014	0.007
Best Fitness Value based on PSO [6]	0.062	0.010	0.004	0.003
Best Fitness Value in the Proposed GA	0.086	0.018	0.006	0.002

Conclusion

In this paper, an unsupervised feature selection method based on genetic algorithm for classification of sound signals is proposed. Unlike the most previous feature selection methods, prior knowledge is not necessary in this case. By using genetic algorithm, the best features are selected. This paper uses a classifier based on nearest neighbor to calculate individual’s fitness value. Each individual over different generations is evaluated by its fitness value. The results of the evaluation illustrates that this new proposed method is very effective for sound classification. By some experiments which are mentioned in experimental results, it can be concluded that the error of classification is decreased by using more features. The same classifier based on MFCC feature selection method is used in order to classify sound signals. By means of MFCC features, nearest neighbor classifier gets the best 99.3% accuracy. On the other hand, with the proposed genetic feature selection algorithm, the same classifier achieves the best 99.8% accuracy. It can be concluded that the proposed feature selection method is more accurate in comparison with the typical audio feature selection methods like MFCC.

References

[1] A. K. Jain and J. Mao, “Guest editorial: Special issue on artificial neural networks and statistical pattern recognition,” IEEE Trans. Neural Networks, vol. 8, no. 1, pp. 1–3, 1997.

[2] B. Bischl, I. Vatolkin, and M. Preuss, “Selecting Small Audio Feature Sets in Music Classification by Means of Asymmetric Mutation,” In Proc. 11th Int. Conf. on Parallel Problem Solving from Nature (PPSN), Krakow, Springer-Verlag, 2010.

[3] D. E. Goldberg, “Genetic algorithms in search, optimization and machine learning,” Addison-Wesley, 1989.

[4] *D. Whitley, “A genetic algorithm tutorial,”. Statistics and Computing. 4 (2): 65–85, 1994.*

[5] E. Alexandre, L. Cuadra, M. Rosa, and F. López Ferreras, “Feature Selection for Sound Classification in Hearing Aids Through Restricted Search Driven by Genetic Algorithms,” IEEE Trans. Audio, Speech, and Lanquage Processing, VOL. 15, NO. 8, 2007.

[6] E. Iranmehr, S. Bagheri Shouraki, and M. M. Faraji, “Unsupervised Feature Selection for Phoneme Sound Classification using Particle Swarm Optimization,” 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2017.

[7] H. Uguz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," Knowledge-Based Systems, vol. 24, no. 7, pp. 1024-1032, 2011.

[8] I. Vatolkin, M. Preuss, and G. Rudolph, “Multi-objective feature selection in music genre and style recognition tasks,” In Proc. of the 13th Annual Conf. on Genetic and Evolutionary Computation (GECCO '11), ACM, New York, USA, 2011.

[9] J. O. García and C. A. R. García, “Mel-Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry with Feed-forward Neural Networks,” in Proceedings of the International Joint Conference, pp. 3140 - 3145, 2003.

[10] M. Büchler, S. Allegro, S. Launer, and N. Dillier, “Sound classification in hearing aids inspired by auditory scene analysis,” EURASIP J. Appl. Signal Process., vol. 18, pp. 2991–3002, 2005.

[11] M. S. Mohamad, S. Deris, S. Yatim, and M. Othman, “Feature selection method using geneticalgorithm for the classification of small and high dimension data,” Proceedings of the 1st International Symposium on Information and Communication Technology, pp. 1-4, 2004.

[12] P. Mermelstein, “Distance measures for speech recognition, psychological and instrumental,” in Pattern Recognition and Artificial Intelligence, C. H. Chen, Ed., pp. 374–388, 1976.

[13] P. Pudil, J. Novovicová, and J. Kittler, “Floating search methods in feature selection,” Pattern Recognit. Lett., vol. 15, pp. 1119–1125, 1994.

[14] R. Leardi, “Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection,” J. Chemometrics, vol. 8, pp. 65–79, 1994.

[15] S. Davis, and P. Mermelstein. “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences,” In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366, 1980.

[16] S. P. Hoseini Alinodehi, and et al, “High-Speed General Purpose Genetic Algorithm Processor,” IEEE Trans. Cybernetics, vol. 46, no. 7, 2016.

[17] X. Huang, A. Acero, and H. Hon. “Spoken Language Processing: A guide to theory, algorithm, and system development,” Prentice Hall, 2001.

[18] Y. Chtioui, D. Bertrand, and D. Barba, “Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision,” J. Sci. Food Agric., vol. 76, pp. 77–86, 1998.

Authors

Mohammad Mahdi Faraji received the B.Sc. degree in electrical engineering and the M.Sc. degree in digital electronics from Amirkabir University of Technology, Tehran, Iran, in 2010 and 2013, respectively, and he is currently Ph.D. student in digital electronics of Sharif University of Technology, Tehran, Iran. He works currently on Data fusion using Fuzzy Logic. His research interests include FPGA, Microprocessor, Advanced Computer Architectures, Sound Localozation, Soft Computing, Neural Netwoks and Machine Learning Algorithms.

email: mmfaraji@ee.sharif.edu

Prof. Saeed Bagheri Shouraki received the B.Sc. degree in electrical engineering and the M.Sc. degree in digital electronics from Sharif University of Technology, Tehran, Iran, in 1985 and 1987, respectively, and the Ph.D. degree in fuzzy control systems from the University of Electro-Communications, Tokyo, Japan, in 2000.
After his Master's studies, he became a faculty member with the Department of Computer Engineering Sharif University of Technology, and continued his activities until 2008. He is currently a Full Professor with the Department of Electrical Engineering, Sharif University of Technology. His research interests include control, robotics, artificial life, and soft computing.

email: s-bagheri@sharif.ir

Ensieh Iranmehr received the B.Sc. degree in electrical engineering from Shahed University, Tehran, Iran, in 2010 and the M.Sc. degree in digital electronics from Amirkabir University of Technology, Tehran, Iran, in 2013. She is currently Ph.D. student in digital electronics of Sharif University of Technology, Tehran, Iran. She works currently on Learning in Spiking Neural Networks. Her research interests include Neural Networks, Fuzzy Systems, Machine Learning Algorithms, Brain Simulation, Soft Computing, FPGA, GPU, and Image Processing.

email: eiranmehr@ee.sharif.edu