International Journal of Systems Signal Control and Engineering Application

MEDWELL PUBLICATIONS SINCE 2000

GA Algorithm Optimizing SVM Multi-Class Kernel Parameters Applied in Arabic Speech Recognition

Aymen Mnassri, Adnane Cherif and Mohammed Bennasr

Laboratory of Analysis and Processing of Signals and Electric and Energy System, FST, Tunis, Tunisia

Key words: Automatic speech recognition, supports vector machines, mel frequency cepstrum coefficients, genetic algorithm

Corresponding Author:
Aymen Mnassri
Laboratory of Analysis and Processing of Signals and Electric and Energy System, FST, Tunis, Tunisia

Page No.: 85-92
Volume: 12, Issue 4, 2019
ISSN: 1997-5422
International Journal of Systems Signal Control and Engineering Application

Abstract: In order to improve the accuracy of Arabic speech recognition, this study proposes a novel recognition technique (ASR) based on GA optimized SVM multi-class algorithm. The Kernel parameters of support vector machine are very important problems that have a great influence on the performance of recognition rate. Thus, GA is adapted to optimize the penalty parameter C and the kernel parameter γ for SVM multi-class which leads to improved classification performance. Finally, the proposed model is tested experimentally using eleven Arabic words mono-locutor. Each word of them is improved by Mel Frequency Cepstral Coefficients (MFCCs) and used as an input to the SVM multi-class classifier. The proposed method enhances the recognition rate which is performed to 100% within short duration training time.

INTRODUCTION

The speech recognition is almost a contemporary discipline of computing which has been in the field of research since 1950's. Speech recognition is an important tool to facilitate the human-machine communication. Thus, with the advancement of automatic speech recognition, the complexity of the integration and recognition problem is increasing. The current speech recognition systems are limited in term of robustness and adaptability to different environments.

The development of automatic speech recognition becomes an interesting domain of research. Hence, the literature is enriched by many researches which treat several methods for speech recognition attending promising results. In fact, several methods have been developed to recognize and classify the speech signal. The most applied methods for speech recognizing are Hidden Markov Models (HMM) (Juang and Rabiner, 1991; Douglas, 2003; O'Shaugnessy, 2008), Multi-Layer Perceptron (MLP) (Ahad *et al*., 2002; Sivaram and Hermansky, 2011) and Self-Organising Maps (SOM) (Venkateswarlu and Kumari, 2011).

All recognition methods have their advantages and inconveniences. Despite its good discriminating ability, the Multi-Layer Perceptron (MLP) has an over training and a local minima problems (Solera-Urena *et al*., 2007). Although, the Self-Organising Maps (SOM) algorithm can easily adapt the new added sample, it is not well defined mathematically. Consequently, the network parameters values need to be established by trial-and-error. So, the ordered mapping, obtained after the training phase, may be missing when used in real environments due to frequent adaptations (Sayers, 1991). Even though the HMM algorithm is the most commonly effective approach used for the recognition stage of an ASR system. However, this method suffers from serious limitations. It is based on the assumption that the probability of being in a particular state is dependent only on its preceding state, ignoring any long-term dependencies, the emission probabilities are arbitrarily chosen as a consequence, these might not even represent correctly the output probabilities of the corresponding state (Trentin and Gori, 2003). The Support Vector Machine models (SVM) have interesting properties in speech recognition such as adaptability, ease of classification of non-linearly separable given noise resistance (Pal and Mather, 2005), good generalization capacity (Ancona *et al*., 2006) and less training set size limitation (Chi *et al*., 2008). Recently, SVMs show their strong classification capabilities, proving to be better than MLP (Solera-Urena *et al*., 2007) but most important research prove also that SVMs can achieve, either comparable or even superior results than the HMMs. But notably it is hard to make choices of SVM kernel function and its parameters. An important factor that affects the performance of SVM is the selection of kernel parameters. Vapnik pointed out that the kernel function parameter and the error penalty parameter C are important factors that influence the performance of SVM (Yuan and Liu, 2007). So, the effectiveness of SVM is determined by parameters (C, γ). Indeed, the selection of the best combination (C, γ) becomes a most important issue that improves the SVM performances.

Compared to the previous cited methods, the main contribution expected by this work is to fields a novel technique based on GA optimized SVM multi-class parameters algorithm, is devoted for Arabic ASR system which can bring several enhancements as:

- The application of the SVM multi- class optimized by GA with the basis of Mel Frequency Cepstral Coefficients (MFCCs)
- The improvement of the recognition rate which achieved 100%
- The reduction of simulation time which constitutes an important criterion for qualifying the system performances

Feature extraction: The feature extraction is the main object of the speech analysis and it is the obligatory passage of all the applications in speech processing, it is a necessity for the next steps such as the recognition. One of the objectives of this analysis is to obtain a compact and informative signal representation. The aim of this step is to propose a simpler representation in the form of an acoustic parameter vector in order to facilitate the extraction of the desired information and to associate with the signal a set of generally acoustic or spectral parameter vectors. The speech signal is a redundant, non-stationary signal but can be considered locally stationary. The analysis of the speech signal takes place during these

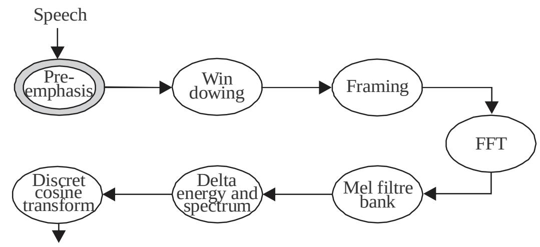

Fig. 1: Calculation of the coefficients MFCC

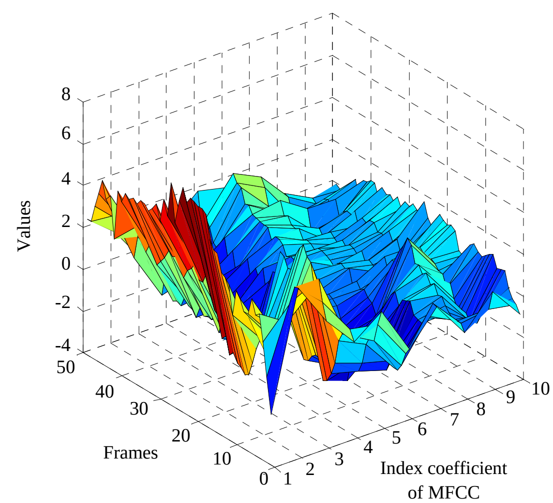

Fig. 2: 3D plots for the results of MFCC

stationary periods, the duration of which varies from 10-30 msec. This duration also corresponds to the stability time of the production model.

The choice of the technical analysis of speech signals is based on three criteria: compactness, robustness and relevance. The most commonly used feature extraction in the speech recognition systems is Mel Frequency Cepstrum Coefficients (MFCCs) (O'Shaughnesssy, 2008) (Fig. 1).

The MFCC (Mel-Scaled Frequency Cepstral Coefficients) calculation principle is derived from psychoacoustic research on the tone and perception of different frequency bands by the human ear (Fig. 2). The FFT passes through a filter bank on the Mel scale. This nonlinear scale mainly takes into account the fact that the perception of the intervals changes according to the area of the spectrum to which the heights composing them belong. The main interest of these coefficients is to extract relevant information in a limited number by relying on both production (Cepstral theory) and speech perception (Mels scale). The calculation proceeds as follows:

- The FFT is calculated on the frames
- The latter is filtered by a bank of triangular filters distributed along the Mel scale. The frequency of the Mel scale is defined by:

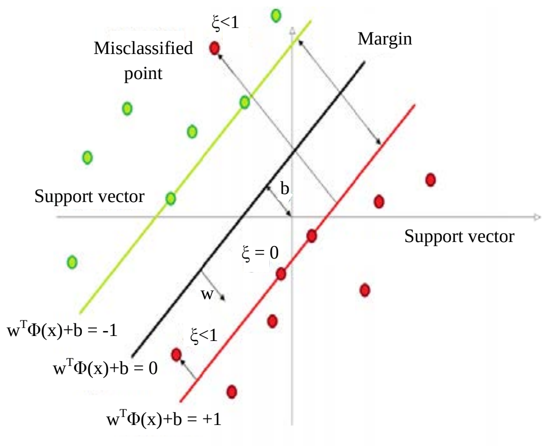

Fig. 3: Binary SVM classification

$$f_{mel} = 2595\times\log\left(1+\frac{f_x}{700}\right) \quad (1)$$

Where:
f_x = The frequency (Hz)
f_{mel} = The Mel-scale frequency of f_x

The logarithm modulus of the output energy of the filter bank is calculated. A reverse discrete cosine transform, (equivalent to the inverse FFT for a real signal) is applied. Finally, to obtain the MFCC coefficients:

$$C_k = \sum_{i=1}^{E} \log E_i \left[\frac{\pi k}{F}(i-\frac{1}{2})\right] \quad (2)$$

Support vector machine: SVM is a classification method for static learning; this latest method has emerged with the theories of Boser *et al*. (1992) which has their efficiency in many applications and is an innovative method in the classification field in the statistical learning like the MLP. SVM is a set of supervised learning techniques and setting its parameters is semi-manually done. The idea of SVM is to find a hyperplane that best separates two classes (Fig. 3).

Binary SVM classification: The separating hyperplane is represented by the following Eq. 3:

$$\omega\times x_i + b = 0 \quad (3)$$

Given a training sample set of example $\{(x_i, y_i), ..., (x_n, y_n)\}$ that can be classified linearly with x_i is the input space and $y_i\in\{-1, 1\}$ is the sample class label, the hyperplane chosen should maximize the distance between the nearest points of each class while remaining a separator. That is to minimize 1/2 $||\omega||^2$ under constraints:

$$y_i\left(\omega\times x_i + b\right) \geq 1 i \in \{1, m\} \quad (4)$$

This is typically solved by the Lagrange multiplier method or the Lagrange is given by:

$$L(\omega, b, \lambda) = \frac{1}{2}\omega\times\omega - \sum_{i=1}^{m}\lambda_i[y_i(\omega\times x_i + b)-1)] \quad (5)$$

where, the coefficients λ_i are the Lagrange multipliers. The Lagrange must be minimized with respect to w and b and maximized in the coefficients λ_i. In case of non-linearly separable training sample set, it is equivalent to minimizing the following quantity:

$$\frac{1}{2}(\omega\times\omega) + c\sum_{I=1}^{m}\xi_i \quad (6)$$

Under the constraint:

$$\omega\times x_i + b \geq 1-\xi_i \text{if} y_i = +1 \quad (7)$$

$$\omega\times x_i + b \geq -1+\xi_i \text{if} y_i = -1 \quad (8)$$

where, $\xi = (\xi_i, ... , \xi_m)$ is slack variable, it controls the further processing of outliers, called "Soft-margin SVM" which controls the extent of punishment to the wrong sub-sample.

Identifying such a nonlinear function is very difficult. the basic idea of support vector machine is: those training set are mapped into a higher-dimensional linear feature space using the kernel function. Where those training set becomes linearly separable in this Features space. This transformation space using a function as follows:

$$F = \{\varphi(x)|x \in X\} \quad (9)$$

Finally, we can obtain the decision function:

$$F(x) = \text{sign}\left\{\sum_{i=1}^{L} a_i y_i k(x_i, x) + b\right\} \quad (10)$$

Where:
a_i = The Lagrange factor get classification results
$k(x_i, x)$ = The kernel function

Many kernel functions that currently used are Smits and Jordan (2002):

- Polynomial kernel function $K_{pol}(x_i, x) = [(x_i, x)+1]^q$
- Gaussian kernel function $K_{rbf}(x_i, x) = \exp(\gamma||x_i, x||^2)$
- Sigmoid kernel function $K_s(x_i, x) = \tanh(g(x_i, x+c)$

The SVM is a new machine learning method based on two classes for the classification of train set, however, it is possible to switch from the binary SVM to the multi-class SVM method that reduce the multi-class problem to a several Bi-class hyperplanes composition allowing to plot the decision boundaries between the different classes. These methods decompose the set of examples into several subsets, each representing a binary classification problem. For each problem, a separation hyperplane is determined by the binary SVM method. In the literature, there are two approaches for decomposition: the "one- against-one" approach proposed by Clarkson and Brown constructs k (k-1)/2 classifiers where each is learned on the data of two classes. "One-against-all" proposed by Scholkopf *et al*. (1999) and al uses a single machine for each group in which each group is formed separately from the rest of the set.

Genetic algorithm: Genetic Algorithms (GA) represent a rather rich and interesting family of stochastic optimization algorithms based on the mechanisms of natural selection and genetics. The fields of application are very diverse.

The basic principles of GAs were developed by Holland (1975). They were inspired by the natural selection mechanism where the best candidates are probably the best adapted to the conditions of competition. The GA then uses a direct analogy with natural evolution. Through the method of genetic evolution, an optimal solution can be found and represented by the last winner of the genetic technique. These algorithms are simple and very efficient in the search for an optimal solution.

GAs function with a population grouping together a set of individuals called chromosomes. Each chromosome consists of a set of genes. For each individual one assigns a calculated value by a function called adaptation function or fitness. In practice, from a population, chromosomes are generated in a random manner during initialization. To define the size of the population, Man etc., mentioned that this size varies from one problem to another. In each cycle of genetic operations, a new population called generation is created from the chromosomes of the current population. For this purpose, certain chromosomes called 'parents' are selected in order to elaborate the genetic operations. The genes of these parents are mixed and recombined for the production of other chromosomes called 'children' constituting the new generation. The steps of the GA are repeated during t cycles; the stopping of the algorithm is fixed according to a stop criterion (Fig. 4). The different steps of GA algorithm are as follows:

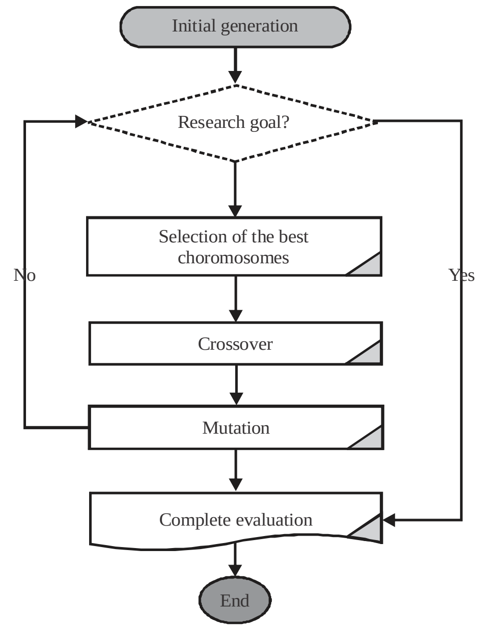

Fig. 4: The flow chart of genetic algorithm

- Step 1: Generate initial population of candidate solution
- Step 2: Find the fitness of each solution
- Step 3: Rank the solutions in terms of their fitness level
- Step 4: Keep more fit solutions and discard the less fit ones
- Step 5: Select and arrange the more fit solutions in pairs for cross over and mutation
- Step 6: Conduct cross over and mutation to give birth to a new generation of candidate solutions
- Step 7: Repeat steps 2-6 until stopping criteria is reached

MATERIALS AND METHODS

The proposed model: In this study, a new model is proposed to recognize Arabic speech based on GA optimizing SVM multi-class kernel parameters, Fig. 5 shows the flowchart of the genetic algorithm method applied to determine the optimal SVM multi-class parameters and to improve the recognition performance.

To further explain our developed approach, we begin by passing our training set through the Mel Frequency Cepstral Coefficients (MFCCs) where each word from training data base is filtered and then windowed by hamming window, then the FFT is applied on each of the windowed word. The signal is then passed through Mel-filter to obtain 12 cepstral coefficients. Finally, the obtained cepstral coefficients are then concatenated to

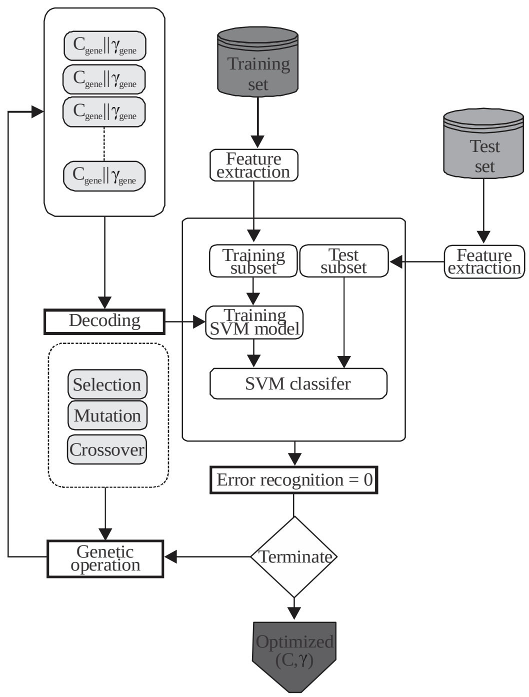

Fig. 5: The flow chart of the improved GA-SVM algorithm

construct one input for SVM classifier. For the classification of train set, we use the "one against all" approach (Scholkopf *et al*., 1999) and our choice is pointed on the RBF kernel because of its higher-dimensional train set classification. The adjustment of the SVM algorithm is decided by two major RBF parameters (C, γ) because of their direct effect on the training results[], the parameter γ is given correspond to a data subspace that has certain dimension and the error penalty parameter C controls the complexity of model and approximates error, obtaining the best performance of SVM classifier is linked to selecting the parameters (C, γ) which is a serious problem on how to choose effectively the best combination (C, γ) to make the performance of SVM reach to its best. Our proposed solution for the problem above concerning the parameters choice is the use of GA optimizing SVM parameter to improve classification performance.

SVM parameters optimization based on GA is realized as follows: Input vectors are speech feature data set and Output vectors are optimal C and γ.

Step 1: Creating a random initial population. This initial population is composed by N chromosome arbitrary representing the SVM parameters (C, γ), each chromosome of those is encoded a binary string, C composed with 3 bytes between 1 and 20 et γ composed with 3 bytes between 1 et 100.

Step 2: Convert the binary chromosome into parameters representing the real value (C, γ).

Step 3: For each chromosomes of the population representing (C, γ), training dataset is used to train the SVM classifier. This classification can be expressed as:

$$\begin{cases} \min\left(\frac{1}{2}(\omega\times\omega)+c\sum_{I=1}^{m}\xi_i\right) \\ K_{rbf}\left(x_i\times x\right) = \exp(-\gamma\|x_i\times x\|^2) \end{cases}$$

The testing dataset is used to verify the prediction performance. This prediction performance is evaluated by the fitness function in our case the objective is the minimization of the prediction error. Each chromosome evaluated by fitness function:

$$\text{Fitness} = 100\text{-recognition rate}$$

Step 4: The stop criterion is either Maximum generation number or fitness = 0, if one of those two criteria is achieved, then the iteration process stops and select the optimal parameters. Otherwise we proceed with the next generation.

Step 5: Generation of a new population, in this work we opted for the "Selection by tournament" method which will select the best N/2 individuals of the initial population according to the value of its function of fitness. Then, we apply genetic operation selection on all individuals including crossover mutation to generate a new population.

Crossover: As the intermediate population is composed of N/2 individuals. We chose to cross the pairs of chromosomes randomly according to the generations by the technique of crossing at a point.

Mutation: Once the new population has reached its desired maximum size, N chromosomes, we try to ensure that our algorithm is able to reach all the points of the search space. This is done by random mutations on the bits of the chromosomes of this population. In this work, we used a Flip Bit in which a mutation operator that simply inverts the value of the chosen gene (0 goes to 1 and 1 goes to 0). If the creation of a new generation is completed, go to Step 2.

RESULTS AND DISCUSSION

This study is devoted to evaluate the classification accuracy of the proposed system in different classification task. Hence, numerous experimental results issued from the test of developed speech recognition approach are

Table 1: The proposed speech recognition system

Parameters	Values
Coefficient MFCC	12
SVM method	SVM multi-class
Approach classify	One against all
Kernel type	RBF
GA coding	Binary
Size of chromosome	20
Scales c	(1.20)
Scales γ	(1.100)
Fitness Function	100-recognition rate
Selection Technique	Selection by tournament
Probability of selection	1/2
Method of mutation	Random
Mutation probability	1/size of chromosome
Method of crossing	One point crossover
Probability of crossover	P cross = 0.5
Stopping criterion	Maximum number of generations or fitness =0

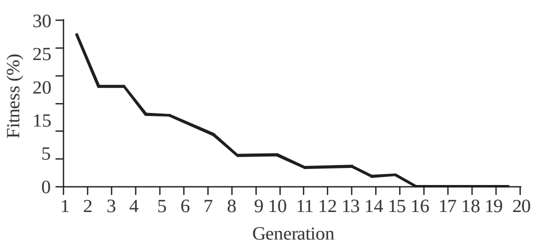

Fig. 6: Iteration process of the GA for optimization SVM parameters

performed. Thus, a database of eleven isolated Arabic recorded by a mono-locutor using a male voice is used. Each word is characterized by a specific duration while the other characteristics as sampling frequency (8 kHz) and the number of used channel (mono channel) are gathered the same for all the recorded words which are repeated 10 times during simulation, the database is divided into two equal parts. The first is used for training while the other is kept for the test. The detail setting parameters used for the proposed speech recognition system is given by Table 1. Figure 6 depicts a typical evolving process of the GA.

Figure 6 shows the different variation of the best fitness of each generation. The stability is obtained when the generation number reaches 16, the maximum fitness value is obtained and stays the same (0%) until the 20th generation which happen in a small amount of time so the GA algorithm is the best choice in determining the best SVM parameters (C, γ), that leads us to 100% recognition accuracy in clean environment.

The recognition experiments are also performed using clean and noisy testing data. Different various noisy conditions, taken from Noisex-92 database: F16 cockpit noise, White Gaussian noise, Rose noise and Volvo car noise with a noise ratio (SNR) from -5-25 dB. The performance evaluation of the proposed model is

Table 2: Performance comparison of recognition accuracy and training time for different methods in clean environment

Speech recognition algorithm	Recognition accuracy	Training time (sec)
GA-SVM	100%	3.284
HMM	89%	82.279
MLP	94	465.580

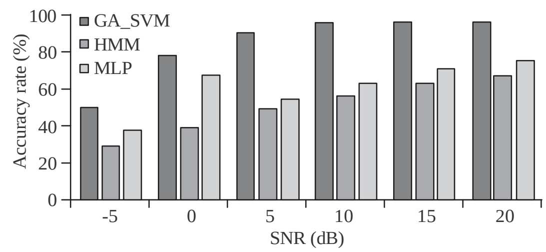

Fig. 7: Comparison of recognition accuracy for GA-SVM, HMM and MLP for white noise

compared with the HMM (Bhara and Kalita, 2015; Alotaibi *et al*., 2010) and MLP (Morgan, 2011; Nasr *et al*., 2012) based speech classification system algorithms without using any speech enhancement algorithm. In all the three algorithms, MFCC is used for the feature extraction.

The performance of a speech recognition system can be measured in terms of accuracy and training time. The recognition accuracy is defined as:

$$\frac{\text{Correctly recognized samples}}{\text{Total number of test samples}}$$

The recognition accuracy and the training time for each technique are given by Table 2. The results of the three speech recognition algorithms are obtained after different tests made in clean environment. According to depicted data, it clearly noted that the GA-SVM seems to be better than the MLP and HMM algorithms. Hence, the proposed algorithm gives the best recognition accuracy in shorter period of training time.

The performance comparison between the three speech recognition algorithms in noisy environment is presented by both Table 3 and Fig. 7-10. The results prove that by applying the proposed GA-SVM model the recognition rate is improved. Indeed, under all noise conditions with different SNRs, the difference between the recognition rates is observed. it can reach 11.82% compared to MLP in case -5 dB with White noise and 36.37% compared to HMM in case -5 dB with Volvo noise. Also, the results obtained show the great capacity of our proposed technical to treat the noisy data with a shorter training time in comparison to HMM and MLP.

Table 3: Performance comparison of recognition accuracy for different methods in noisy environment

Noisy types	Speech recognition algorithm	-5 dB	0 dB	5 dB	10 dB	15 dB	20 dB
White	GA-SVM	50.90	78.18	90.90	96.36	96.36	96.36
	HMM	29.09	40	49.09	56. 36	63.63	67.27
	MLP	38.18	67.27	54.54	63.63	70.90	76.36
F16	GA-SVM	32.72	69.09	87.27	96.36	96.36	96.36
	HMM	25.45	36,36	56.36	56.36	63.63	70.90
	MLP	27.27	34.54	63.63	80	83.63	81.81
Rose	GA-SVM	69.09	92.72	96.36	98.18	100	100
	HMM	27. 27	29.09	40	49. 09	58.18	72.72
	MLP	27.27	61.81	63.63	69.09	72.72	80
Volvo	GA-SVM	89.09	96.36	98.18	98.18	98.18	98.18
	HMM	52.72	61.8	72.72	74.54	76.36	80
	MLP	69.09	69.09	72.72	78.18	80	83.63

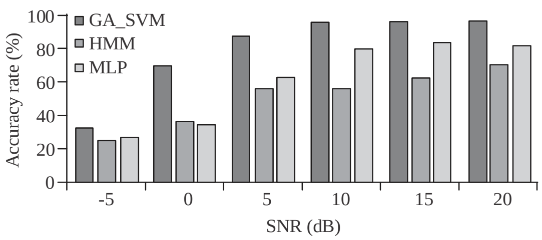

Fig. 8: Comparison of recognition accuracy for GA-SVM, HMM and MLP for F16 noise

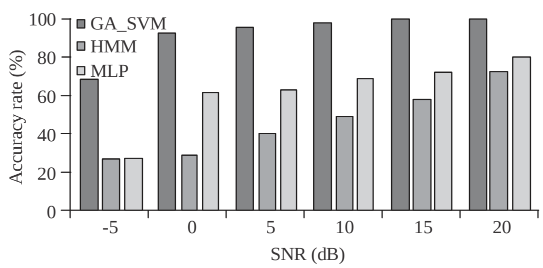

Fig. 9: Comparison of recognition accuracy for GA-SVM, HMM and MLP for rose noise

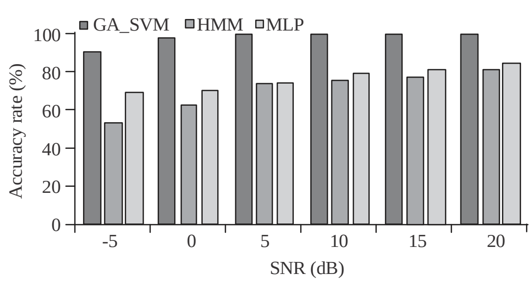

Fig. 10: Comparison of recognition accuracy for GA-SVM, HMM and MLP for volvo noise

CONCLUSION

In this study, a new technique for Arabic speech recognition using the GA optimizing SVM multi-class kernel parameters has been presented. The obtained results of the proposed method prove that GA is an effective solution to optimize SVM parameters; it can improve the learning ability of SVM that leads us to 100% recognition accuracy in clean environment. Moreover, the evaluation of the proposed method is performed by comparing it to the speech recognition approach based on HMM and MLP using clean and noisy testing data without using any speech enhancement algorithm. This evaluation which is based on terms of precision and speed show that the proposed technique provides, is better performance than the existing technique like HMM and MLP based speech recognition techniques.

RECOMMENDATION

In future research and with these encouraging results, we aspire to develop an embedded system with our proposed method.

REFERENCES

Ahad, A., A. Fayyaz and T. Mehmood, 2002. Speech recognition using multilayer perceptron. Proceedings of the IEEE Students Conference ISCON’02, Vol. 1, August 16-17, 2002, IEEE, Lahore, Pakistan, pp: 103-109.

Alotaibi, Y.A., M. Alghamdi and F. Alotaiby, 2010. Speech recognition system of Arabic alphabet based on a telephony Arabic corpus. Proceedings of the International Conference on Image and Signal Processing, June 30-July 2, 2010, Springer, Trois-Rivieres, Canada, pp: 122-129.

Ancona, N., R. Maglietta and E. Stella, 2006. Data representations and generalization error in kernel based learning machines. Pattern Recognit., 39: 1588-1603.

Bharali, S.S. and S.K. Kalita, 2015. A comparative study of different features for isolated spoken word recognition using HMM with reference to Assamese language. Int. J. Speech Technol., 18: 673-684.

Boser, B.E., I.M. Guyon and V.N. Vapnik, 1992. A training algorithm for optimal margin classifiers. Proceedings of the 5th Annual Workshop on Computational Learning Theory, July 27-29, 1992, ACM, Pittsburgh, Pennsylvania, pp: 144-152.

Chi, M., R. Feng and L. Bruzzone, 2008. Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. Adv. Space Res., 41: 1793-1799.

Douglas, O., 2003. Interacting With computers by voice: Automatic speech recognition and synthesis. Proc. IEEE, 9: 1272-1305.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. 1st Edn., University of Michigan Press, Ann Arbor, MI., USA., ISBN-13: 9780472084609, Pages: 183.

Morgan, N., 2011. Deep and wide: Multiple layers in automatic speech recognition. IEEE Trans. Audio Speech Lang. Process., 20: 7-13.

Nasr, M.B., M. Talbi and A. Cherif, 2012. Arabic speech recognition by bionic wavelet transform and MFCC using a multi layer perceptron. Proceedings of the 2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), March 21-24, 2012, IEEE, Sousse, Tunisia, pp: 803-808.

O'Shaughnessy, D., 2008. Automatic speech recognition: History, methods and challenges. Pattern Recognit., 41: 2965-2979.

Pal, M. and P.M. Mather, 2005. Support vector machines for classification in remote sensing. Intl. J. Remote Sens., 26: 1007-1011.

Rabinar, J., 1991. Hidden Markov models for speech recognition. J. Technometices, 338: 251-272.

Sayers, C., 1991. Self organising feature maps and their applications to robotics. Report No.: MS-CIS-91 46, Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Philadelphia.

Scholkopf, B., O. Simard, A. Smola and V. Vapnik, 1999. Prior Knowledge in Support Vector Kernels. Vol. 2, The MIT Press, Cambridge, Massachusetts,.

Sivaram, G.S. and H. Hermansky, 2011. Sparse multilayer perceptron for phoneme recognition. IEEE. Trans. Audio Speech Lang. Process., 20: 23-29.

Smits, G.F. and E.M. Jordaan, 2002. Improved SVM regression using mixtures of kernels. Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02 (Cat. No. 02CH37290) Vol. 3, May 12-17, 2002, IEEE, Honolulu, Hawaii, pp: 2785-2790.

Solera-Urena, R., J. Padrell-Sendra, D. Martin-Iglesias, A. Gallardo-Antolin, C. Pelaez-Moreno and F. Diaz-de-Maria, 2007. SVMS for Automatic Speech Recognition: A Survey. In: Progress in Nonlinear Speech Processing, Stylianou Y., M. Faundez-Zanuy and A. Esposito (Eds.)., Springer, Berlin, Germany, pp: 190-216.

Trentin, E. and M. Gori, 2003. Robust combination of neural networks and hidden Markov models for speech recognition. Proc. IEEE, 6: 1519-1531.

Venkateswarlu, R.L.K. and R.V. Kumari, 2011. Novel approach for speech recognition by using self-organized maps. Proceedings of the 2011 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), April 22-24. 2011, IEEE, Udaipur, India, pp: 215-222.

Yuan, X.Y. and A.L. Liu, 2007. Kernel parameter selection of the support vector machine based on .particle swarm optimization. Control Theory Appl., 26: 5-8.