Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (22): 5349-5355, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.5349.5355

Practical Speech Emotion Recognition Based on Im-BFOA

[1]Bao Yongqiang, [2]Xi Ji and [2]Xu Haiyan
[1]School of Communication Engineering, Nanjing Institute of Technology, 211167, Nanjing, China
[2]School of IOT Engineering, Hohai University, 213022, Changzhou, China

Abstract: For Support Vector Machines (SVM) parameter optimization problem, we propose an improved bacterial foraging algorithm (Im-BFOA) and increase its learning ability in the practical speech emotion recognition. Firstly, we introduce simulated annealing (SA),Gaussian mutation and chaotic disturbance operator into BFOA to balance the efficiency of search and the diversity of population. Secondly, use Im-BFOA to optimize SVM parameters and propose a Im-BFOA-SVM method; Thirdly, based on prosodic features, quality features and chaotic features of speech, build a 144-dimension emotional feature vector and use FDR to dimension reduction to 5 dimensions; Finally, test the algorithm performance on the practical speech emotion database and compare the proposed algorithm Particle Swarm Optimization(PSO) algorithm to optimize the parameters of SVM (PSO-SVM method) with basic SVM methods and Back-Propagation (BP) neural network method. Experimental results show that average recognition rate of the Im-BFOA-SVM method reached 78.1%, respectively, higher than PSO-SVM method, SVM methods and BP neural network method of 3.7, 5.4 and 9.8%, indicating that Im-BFOA is a kind of effective SVM parameter selection method which can significantly improve practical speech emotion recognition rate.

Key words: Im-BFOA, SVM, speech emotion recognition, neural network

INTRODUCTION

In the field of speech emotion research, current researches focus on the basic emotion recognition, such as happiness, sadness, anger and calm (Yeh *et al*., 2011; Zhao *et al*., 2008). However, in practical applications, there exist irritability and other practical speech emotions, so limited to the research of recognition of the basic speech emotions cannot meet the needs of practical application. There are several common methods of speech emotion recognition in the earlier time, such as neural network (Nicholson *et al*., 2000), Hidden Markov Models (HMM) (Nwe *et al*., 2003), Gaussian Mixture Model (GMM) (Huang *et al*., 2011). Neural network method was an earlier method used for speech emotion recognition. It can obtain the difficulty level of the recognition of positive and negative emotions through analyzing the basic emotion categories. However, the recognition rate of neural network method is low and the average recognition rate is only 50%. As a result of using short-term timing characteristics, HMM methods are easily affected by the changes of text messages, for example, formant is a common speech emotion characteristic but seriously affected by the phoneme information. In recent years, GMM method is a relatively successful method in speaker and language recognition but it strongly depends on the training data and it=s difficult to set the parameters of mixing degree and it needs to be adjusted during the experiment which leads to the algorithm's low generality in different problems or databases.

In practical applications, speech emotion recognition has some difficulties in collecting enough data and collecting data of special emotion etc., so emotion recognition in small sample conditions is important. Support Vector Machine (SVM) (Cortes and Vapnik, 1995) is widely used in areas like pattern classification because of its high classification accuracy, strong learning ability and good generalization performance. It is feasible to apply SVM to speech emotion recognition in small sample conditions. In the specific design of SVM classifier, classification accuracy strongly depends on the parameter selection. In this study, we use heuristic optimization algorithm to optimize the parameter of SVM. There are many successful applications of heuristic optimization algorithm in recent years' literature, for example, the parameter optimization based on Particle Swarm Optimization (PSO) and genetic algorithms can improve the speed of parameter optimization. However, the PSO algorithm (Shao *et al*., 2006) and GA algorithms (Chen *et al*., 2004) are prone to premature convergence which leads to the low accuracy of parameter optimization.

Corresponding Author: Bao Yongqiang, School of Communication Engineering, Nanjing Institute of Technology, 211167, Nanjing, China Tel: 13813929100

Bacteria Foraging Optimization Algorithm (BFOA) is designed to simulate foraging behavior of human coliform by Passino (2002) in 2002 which causes the researchers in different fields attention to its intuitive construction and intelligible natural mechanism. In order to improve the algorithm tendency and copy operations, coordination processing algorithm of local mining capacity and global exploration capability, many scholars have conducted a targeted research. Mishra (2005) proposed to choose the optimal step length by using fuzzy inference mechanism which was called Fuzzy Bacterial Foraging (FBF). But the performance of FBF was fully depended on the selection of membership functions and fuzzy rules parameters. Except repeated experiments, for a given problem, it didn't have a systematic way to determine the parameters, so this algorithm had no versatility. In the research of copy operations, Abraham *et al*. (2008) and others concluded that it could avoid algorithm precocity by adding adaptive mechanism into copy operations with the theoretical analysis of algorithm convergence and stability. But this theoretical only considered the copy operations of two particles formed population in one-dimensional continuous space built on the assumption of certain conditions which was the same as the theoretical analysis of trends.

Therefore, in this study simulated annealing (Brooks and Morgan, 1995) is introduced into the global information exchange phase to improve the computational efficiency and solution accuracy during global information exchange. Gaussian mutation and chaotic disturbance operator was introduced into partial depth search to improve the ability of jumping out of local extreme, population diversity and convergence speed in the late iteration. The Im-BFOA would be used in the SVM training as the method of SVM parameters optimization which was called Im-BFOA-SVM here. In order to verify the algorithm effectiveness, the emotional acoustic parameters was extracted from a set of actual data. And this algorithm would be used for the first time in then practical speech emotion recognition.

OPTIMIZATION METHOD OF SVM PARAMETERS BASED ON IM-BFOA

Im-BFOA fundamentals: Firstly, drawing lessons from the SA annealing algorithm mechanisms to improve the replication of BFOA algorithm, Gaussian mutation and chaotic disturbance operations are introduced. Specific approach is to do the Gaussian mutation for all individuals which have the energy value superior to the average, if $\min\{1, \exp(-\Delta f / T)\} > \text{rand}()$ (here, Δf is the energy difference between the individual after mutation and before mutation, T is the annealing temperature), then use the individual after Gaussian mutation instead of the before, or the original individual remains unchanged; do the chaotic disturbance for all individuals which have the fitness value worse than the average, if $\min\{1, \exp(-\Delta f/T)\}>\text{rand}()$, then use the individual after chaotic disturbance instead of the before, or the original individual remains unchanged. This action makes late iterative algorithm population diversity become improved, the ability to jump out of local minima become stronger, convergence become faster and eventually converges to the global optimum.

Based on the above ideas, this study constructs a combination of SA with Gaussian mutation and BFOA with chaos disturbance, namely Im-BFOA, the specific optimization iteration steps are:

Step 1: Initialization parameters p, S, N_c, N_S, N_{re}, N_{ed}, P_{ed}, G(i)(i = 1, 2,…, S)

Here, p is the dimension of the search space, ■ is the size of bacterial populations, N_c represents the number of bacteria's approach behavior, N_S represents the maximum number of steps that the operation tendency forward in one direction,, N_{re} represents the number of bacteria's replication behavior, N_{ed} represents the number of bacteria's migratory behavior, P_{ed} represents the probability of migration, C(i) represents the step of swimming forward.

Step 2: Number of operations to initialize migration l, the replication operation times k, the number of tendency operation j

Step 3: Conduct tendency operation:

- Make the bacteria i step towards as following, i = 1, 2,…, S
- Calculate the fitness function J(i, j, k, l) of bacteria i:

$$J(i,j,k,l) = J(i,j,k,l) + J_{cc}(\theta^i(j,k,l), P(j,k,l)) \quad (1)$$

where, J_{CC} is the affection value of signal transmitted between the population of bacteria

- Select the best fitness value of bacteria i: J_{best} = J(i, j, k, l)
- **Rotate:** Generate a random variable Δi, where each element $\Delta_m(i)$, (m = 1, 2,..., p are random numbers distributed in [-1, 1])
- **Move:** Bacterial i randomly generated after the rotation step in the direction of long size swimming C(i), we get:

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}} \quad (2)$$

- Update the fitness value:

$$J(i,j+1,k,l) = J(i,j,k,l) + J_{cc}(\theta^i(j+1,k,l), P(j+1,k,l)) \quad (7)$$

- **Swimming:** If $J(i, j+1, k, l) < J_{best}$, set $J_{best} = J(i, j+1, k, l)$ and update $\theta^i(j+1, k, l)$ according to Eq. 2
- Back to step 2, deal with next bacteria i+1

Step 4: If the number of tendency operation is less than N_C, return to step 3, continue tendency operation

Step 5: Replication

For a given k, l and each i = 1,2,…, S, range the bacterial energy value from small to large. Do Gaussian mutation for all individuals which have the energy value superior to S/2, if $\min\{1, \exp(-\Delta f/T)\} > \text{rand}()$, then use the individual after Gaussian mutation instead of the before, or the original individual remains unchanged; do the chaotic disturbance for all individuals which have the fitness value worse than S/2, if $\min\{1, \exp(-\Delta f / T)\} > \text{rand}()$, then use the individual after chaotic disturbance instead of the before, or the original individual remains unchanged. Remove bacteria whose energy values ranked before S_r = S/2, select bacteria of larger energy values ranked after S_r to replicate, each split into the same bacterium.

Step 6: If the number of replication operations is less than N_{re}, initialize the number of tendency operation $_j$, then return to step 3

Step 7: **Migration:** After several generations' replication operations of Bacterial flora, each bacterium is re-distributed randomly to optimization space by probability P_{ed}. If the temperature reaches the ground state T_g, evolutionary process is completed, then output the global optimum value; Otherwise, modify population's annealing temperature, namely set T = c.T, initialize the number of replication operations $_k$ and tendency operations $_j$, then return to step 3

Optimization of SVM parameters based on Im-BFOA: Among Classical SVM training methods, selection of relevant parameters depends on experience or any given, such selecting SVM parameters randomly for training often results in that they obtained model's classification is not very satisfactory. To obtain SVM parameters based on Cross Validation (CV) can get better classification results than on random selection. Common CV methods: Hold-Out Method, K-fold Cross Validation (K-CV) and Leave-One-Out Cross Validation (LOO-CV).

Although, traditional grid search method can get highest recognition under the sense of CV, it needs to traverse all the parameters within the grid points. If the search scope is very large, then the traditional grid search method will become very difficult and time consuming while the heuristic search method will save a lot of time and can quickly search the global optimum value without traversing all the parameters within the grid points. Therefore, this study uses Im-BFOA to search optimal SVM parameters under the sense of K-CV. The proposed optimization is applied to design the parameters of SVM classifier, the key is to find a reasonable fitness function. Since, do the classification accuracy optimization for training sample set under the sense of K-CV is to find the maximum value, this study chooses classification accuracy's opposite number as the fitness function under the sense of K-CV. Here choose Gaussian radial basis kernel function (Radial Basis Function, RBF) as the kernel function of SVM, so we need to optimize the SVM parameters, namely penalty coefficients C' and RBF width σ, C' and σ constitute a two-dimensional individual bacterium. The use of Im-BFOA to search for the optimal SVM parameters for SVM training, the specific steps are:

- Randomly initialize bacterial populations and chaotic mapping variable initial values, each bacterium individual corresponds to (C', σ). Initialize C' and σ as the random number, respectively on the interval [0, θ] and [0, δ] (θ and δ is a positive number). Set the group number K of K-CV and the parameters of Im-BFOA
- Calculate fitness values of all bacteria individuals. Use the optimization steps of Im-BFOA to optimize bacteria individual
- If the termination condition is satisfied, then output the global optimum individual (C'_g, σ_g), here, C'_g and σ_g are the optimum penalty coefficient and RBF width, optimization process ends; otherwise, go to step 2

After the completion of the parameter optimization, we can build a identification system based on Im-BFOA-SVM, the system block diagram shown in Fig. 1. We use the optimal parameters (C'_g, σ_g) found by Im-BFOA in the SVM training to get the best training model and then use the model for sample identification. Thus we build a Im-BFOA-SVM method.

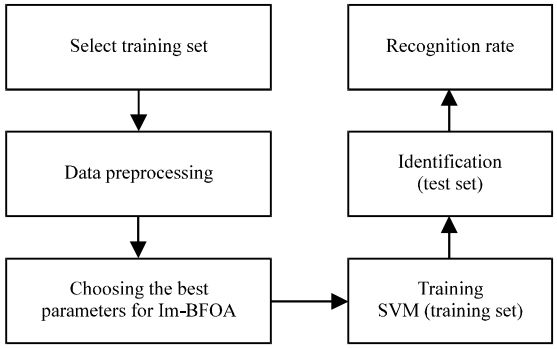

Fig. 1: Identification model based on Im-BFOA-SVM method

APPLICATION OF PRACTICAL SPEECH EMOTION RECOGNITION

Practical speech emotion database: At present, the mainstream speech emotion database (Amir *et al*., 2000; Paeschke and Sendlmeier, 2000) only contains the basic speech emotion category internationally but doesn't involve practical speech emotion such as anxiety category which cannot satisfy the needs of practical application. Therefore, Southeast University laboratory recorded the database containing practical speech emotion in the project cooperation with Fujitsu. The experimental data in this study is all from that database.

In the process of recording the database, the recording software and hardware equipments include: one high-performance computer, one M-audio mobile Pre USB sound card, one M-audio Nova large membrane capacitance microphone, one monitor earphone etc. The recording software Cool Edit pro 2.0 was set by mono, 16-bit sampling precision and 48 KHz sampling frequency. The recorded audio was saved as a PCM coding wav format. The recording process was preceded in a quiet laboratory. Fifty-one college students including 23 men and 28 women whose ages were between twenty and thirty years old were chosen to participate in the recording of the database. They were asked to have clear speaking, standard pronunciation and good command of mandarin, good health and normal hearing but without a recent cold. We choose twenty-five short phrases from no emotion tendentiousness corpus to obtain four target emotions which were irritable, happy, neutral and sad emotion through the way of inducing. The specific induce mode is shown in Table 1.

Ten students who did not participate in the recording work were chosen to take perception experiment to screen out training and testing data with significantly emotional expression and better recording quality. 3707 statements were kept to be used as the training and testing data. Among them, 898 were irritable, 920 were happy, 945 were neutral and 944 were sad.

Table 1: Induce mode of speech emotion

Target emotion	Induce mode	Concrete implementation method
Irritable	Noise induce	Through the headset noise to induce irritable emotion
Happy	Film and TV clips induce	Through the comedy clips to induce positive emotion
Neutral	No need to induce	Read statements without any emotion
Sad	Performance induce	Recall the memories of sad experience to induce

Emotional feature selection: It is difficult to effectively distinguish all kinds of emotions only from several characteristics, so this study extracts the prosodic features, quality features and chaotic characteristics of the emotional statements. Short-time energy, pronunciation duration, speed and pitch frequency are usually considered to be prosodic features which are closely associated with the wake dimension of emotional dimension model. Quality feature mainly refers to voice tone and language spectrum characteristics, so it is also called the segmental feature. From the perspective of speech production model, quality feature mainly refers to spectral envelope characteristic related to channel response. Representative quality features mainly include formant, cepstrum, LPC coefficients and its derivative parameters etc. However, for the emotion of the natural language, the differences on activation dimension are much smaller generally than the show emotion, the traditional prosodic features and quality features could not recognize the changes of true emotions very well, thus, the speech chaotic characteristics which can reflect the potency dimension changing information are of special important value in the practical speech emotion recognition. The chaotic characteristics include correlation dimension, largest Lyapunov index and Kolmogorov entropy parameter. Therefore, this study constructed a 144-dimensional speech emotion feature vector by the method of global statistical characteristic.

EXPERIMENTAL RESULTS

In order to make sure that the Im-BFOA-SVM method has a good effect in practical effect of speech emotion recognition, We do the recognition experiment comparing with PSO-SVM method, SVM method and BP Neural Network method. According to the D. Ververidis' research (Ververidis *et al*., 2004), four or five features are enough to describe the emotional distribution and it can get a better recognition rate. In the experiment, the Fisher

Discriminant Ratio coefficient method descends the 144D original feature vector to 5D, in order to do the emotion recognition experiment which has nothing to do with the text information and the speakers. Then we randomly selected 300 statements as the training sample set (75 for each emotion) from the data of the identify experiment and the rest 3407 is as a test sample set. Before extracting the speech features, the windows for separating the frames about 25ms will be added to the voice signal, the two adjacent frames overlap 1/2 and it uses the hamming window to reduce the truncation effect in each frame edge. In the experiment, the BFOA Algorithm parameters are set as below: the space dimensionality is 100, bacterial population size is 100, the number of trend behavior is 10, the maximum step is 4, the number of copy behavior is 10, the number of migrating behavior is 2, The migrating probability is 0.2, the Moving forward step is 0.2; Fitness function is the opposite number of the classification accuracy rate in K-CV method with the training sample set and the K-CV group number is 3, the local iteration number of the subgroup is 30, initial temperature of annealing is 10, cooling temperature coefficient is 0.9, the base state temperature is 1, inoculation probability is 0.4, population size of the PSO Algorithm is 200, the dimensions of the solution is 2, the iteration number is 100, weighting factor the kernel function of the SVM is Gaussian RBF, the search scope of the punish coefficient C' is [0.1,100], the search scope of the width of the Gaussian RBF is [0.01,1000], the BP neural network hidden layer node number is 15, the activation function is the Sigmoid tansig, output layer function is purelin, the training function is trainlm, training target accuracy is $1x10^{-5}$, training number is 100, the search scope of both of the weight and the threshold is [-1,1]. Due to the proportion of the training data is smaller, GMM mixing degree is also set smaller accordingly. This experiment is set as 8 and it uses the K-means clustering algorithm for initialization, uses the EM algorithm to estimate the parameters of the Gaussian mixture model and the maximum iteration number of EM algorithm is 40; Fig. 2 shows the results of four methods. The first kind of error rate is the total result that this kind of emotional is

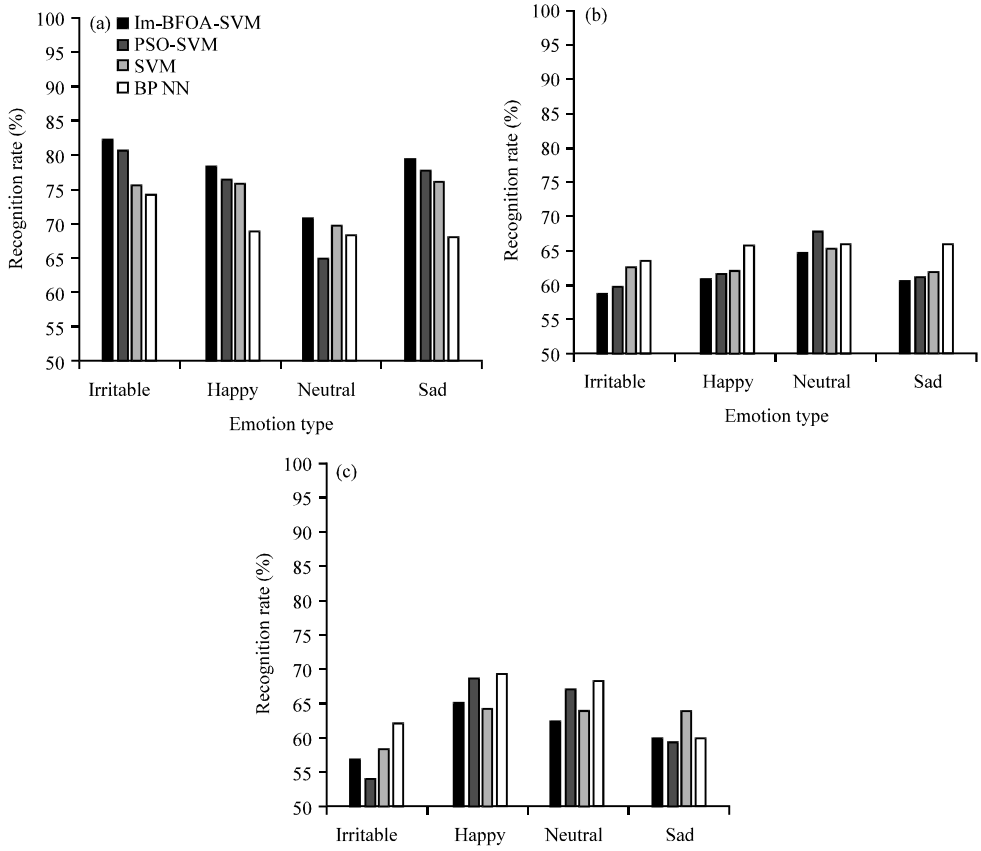

Fig. 2(a-c): Six kinds of recognition methods of recognition results comparison (a) Recognition rate comparison, (b) First kind of error rate comparison and (c) Second kind of error rate comparison

mistaken for the other three kinds of emotional categories. The second kind of error rate is the total result that the other three kinds of emotional categories is mistaken for this kind of emotional.

From the above chart results, we can see:

- Average recognition rate of the PSO-SVM method is 74.4%, The average recognition rate of the SVM method is 72.7%, The average recognition rate of the BP Neural Network method is 68.3% and when using the Im-BFOA-SVM method into the practical speech emotion recognition such as upset, the average recognition rate is up to 78.1% and the recognition rate improves obviously, increased by 3.7, 5.4 and 3.7%, respectively than the PSO-SVM method, the SVM and BP Neural Network. Among them, recognition effect of the upset emotion is the best, up to 82.7% and significantly higher than that of other three methods. As a result, Im-BFOA-SVM methods are suitable for practical speech emotion recognition
- Among the four kinds of emotion considered, the "upset" and "sad" emotion are close in potency dimension coordinates, so misjudgment between them is obvious. However, the discrimination among the "upset", "happy" and "neutral" is better and misjudgment rarely occurs. This is because that they are far away in potency dimension coordinates. So you can get a satisfactory recognition rate as long as you select the proper potency dimension parameters when identifying these emotions

CONCLUSION

This study proposed Im-BFOA method which introduces SA thought in the stage of global information exchange, introduces Gaussian mutation and chaos perturbation in the depth of local search to improve the search efficiency of the algorithm and the ability of getting rid of local extreme value. In order to improve the classification accuracy of SVM, an Im-BFOA-SVM method was proposed by selecting relevant parameters of the SVM using Im-BFOA. Furthermore, good recognition effect is obtained by applying the algorithm into the speech emotion recognition.

ACKNOWLEDGMENTS

This study was supported by China Postdoctoral Science Foundation (No. 2012M520973), the Scientific Research Funds of Nanjing Institute of Technology (No. ZKJ201202) and the underwater acoustic signal processing open research fund of Southeast University of Key Laboratory of Ministry of Education (B) (No. UASP1202).

REFERENCES

Yeh, J.H., T.L. Pao, C.Y. Lin, Y.W. Tsai and Y.T. Chen, 2011. Segment-based emotion recognition from continuous Mandarin Chinese speech. Comput. Hum. Behav., 27: 1545-1552.

Zhao, Y., L. Zhao, C. Zou and Y. Yu, 2008. Speech emotion recognition using modified quadratic discrimination function. J. Electron., 25: 840-844.

Nicholson, J., K. Takahashi and R. Nakatsu, 2000. Emotion recognition in speech using neural networks. Neural Comput. Appl., 9: 290-296.

Nwe, T.L., S.W. Foo and L.C. De Silva, 2003. Speech emotion recognition using hidden Markov models. Speech Commun., 41: 603-623.

Huang, C.W., Y. Zhao, Y. Jin, Y.H. Yu and L. Zhao, 2011. A study on feature analysis and recognition of practical speech emotion. J. Electron. Inform. Technol., 33: 112-116.

Cortes, C. and V. Vapnik, 1995. Support-vector networks. Machine Learn., 20: 273-297.

Shao, X.G., H.Z. Yang and G. Chen, 2006. Parameters selection and application of support vector machines based on particle swarm optimization algorithm. Control Theory Appl., 23: 740-743.

Chen, P.W., J.Y. Wang and H.M. Lee, 2004. Model selection of SVMs using GA approach. Proceedings of the IEEE International Joint Conference on Neural Networks, Volume 3, July 25-29, 2004, Budapest, Hungary, pp: 2035-2040.

Passino, K.M., 2002. Biomimicry of bacterial foraging for distributed optimization and control. IEEE Control Syst., 22: 52-67.

Mishra, S., 2005. A hybrid least square-fuzzy bacterial foraging strategy for harmonic estimation. IEEE Trans. Evol. Comput., 9: 61-73.

Abraham, A., A. Biswas, S. Dasgupta and S. Das, 2008. Analysis of reproduction operator in bacterial foraging optimization algorithm. Proceedings of the IEEE Congress on Evolutionary Computation, June 1-6, 2008, Hong Kong, pp: 1476-1483.

Brooks, S.P. and B.J.T. Morgan, 1995. Optimization using simulated annealing. J. R. Stat. Soc. Ser. D (Statistician), 44: 241-257.

Amir, N., S. Ron and N. Laor, 2000. Analysis of an emotional speech corpus in Hebrew based on objective criteria. Proceedings of the Tutorial and Research Workshop on Speech and Emotion, September 5-7, 2000, Newcastle, Northern Ireland, UK., pp: 29-33.

Paeschke, A. and W.F. Sendlmeier, 2000. Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements. Proceedings of the Tutorial and Research Workshop on Speech and Emotion, September 5-7, 2000, Newcastle, Northern Ireland, UK., pp: 75-80.

Ververidis, D., C. Kotropoulos and I. Pitas, 2004. Automatic emotional speech classification. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Volume 1, May 17-21, 2004, Montreal, Canada, pp: 593-596.