33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Binary Particle Swarm Optimization for Feature Selection in Detection of Infants with Hypothyroidism

A. Zabidi, L. Y. Khuan (*MIEEE-EMBS*), W. Mansor (*MIEEE-EMBS*), I. M. Yassin (*MIEEE-EMBS*), and R. Sahak

***Abstract*—Hypothyroidism in infants is caused by the insufficient production of hormones by the thyroid gland. Due to stress in the chest cavity as a result of the enlarged liver, their cry signals are unique and can be distinguished from the healthy infant cries. This study investigates the effect of feature selection with Binary Particle Swarm Optimization on the performance of MultiLayer Perceptron classifier in discriminating between the healthy infants and infants with hypothyroidism from their cry signals. The feature extraction process was performed on the Mel Frequency Cepstral coefficients. Performance of the MLP classifier was examined by varying the number of coefficients. It was found that the BPSO enhances the classification accuracy while reducing the computation load of the MLP classifier. The highest classification accuracy of 99.65% was achieved for the MLP classifier, with 36 filter banks, 5 hidden nodes and 11 BPS optimised MFC coefficients.**

I. Introduction

Crying is the only communication channel of infants. Hence, cry signals are rich in information, embedding the infant emotional and pathologic conditions. Due to their unique nature, many types of intelligent classifiers have been attempted to extract the information. Of these, the Artificial Neural Network (ANN) method is found to be able to distinguish the infant emotions based on their cry signals with some success [1]. The use of Time-Delay-Neural Network (TDNN) in classifying pathologic states such as normal, deaf and asphyxiated infants, has reported up to 86.06% classification accuracy [2]. The same study has also shown strong correlation between the infant cry signals and diseases [1].

Hypothyroidism occurs in infants born without, or with malfunctioning thyroid glands. Symptoms such as prolonged or recurrent jaundice, delay in umbilical cord separation and umbilical hernia are indicators of such disease in infants [2]. Insufficient production of hormones by the thyroid gland is the cause for hypothyroidism. The enlarged liver increases stress in the chest cavity. This causes the sound of the infant cry to become husky and low, making them unique enough to be distinguished from the healthy infant cries [2].

The automatic infant cry classification process consists of two processing stages; feature extraction and pattern recognition [1, 3-5]. The objective to classify the infant cry is to detect physiologic defects from input waveforms of infant cries. Before classification can be performed, the feature extraction stage is carried out first. The most common feature extraction method used is the Mel Frequency Cepstrum (MFC) analysis. One of the drawbacks of MFC analysis is an overwhelming of coefficients for signal representation, which incurs heavy penalty on the computation cost and efficiency for post-processing. Furthermore, each coefficient has different levels of significance for classification. Therefore, a mechanism to reduce the original dataset by selecting only the significant features that is just sufficient to represent the signal must be included in the classification process.

Binary Particle Swarm Optimization (BPSO) is a swarm-based optimization algorithm that is able to solve binary optimization problems [6]. It achieves this by representing the potential solutions in binary form and the continuous valued particle positions as the probability of a bit switch in the original solution to occur. By altering the probability of the bits to switch, new solutions can be generated across iterations. The BPSO is thus used to select significant features from the MFC analysis. It has been used in our previous work to model heat exchanger with NARX [7].

In this work, the effectiveness of the BPSO to perform the selection of MFC features for classification of hypothyroidal infant cry is investigated. The hypothyroidal infant cries are distinguished from the normal ones with a MultiLayer Perceptron (MLP) ANN. The infant cry features transformed by the MFC analysis serve as the input vector to the MLP classifier. It is hypothesized that pre-selection of significant features using BPSO can improve the overall accuracy of the MLP classifier.

II. Theoretical Background

A. Mel Frequency Cepstral Analysis

MFC analysis is an acoustic feature extraction method that takes human frequency perception sensitivity into consideration during extraction, which makes it ideal for voice recognition [8]. The method extracts MFC coefficients to provide a good representation of the dominant acoustic features in selected time windows. For a signal with N frames, $n = 0, 1, \ldots (N\text{-}1)$, the MFC analysis is defined as:

$$c(n) = DCT\left(log\left(\left|FFT\left(s(n)\right)\right|\right)\right) \tag{1}$$

Manuscript received March 10, 2011. Authors are with the Faculty of Electrical Engineering, University Teknologi Mara, 40450 Shah Alam, Selangor, Malaysia (e-mail: leeyootkhuan@salam.uitm.edu.my)

where $c(n)$ is the MFC coefficients at *nth* frame, $s(n)$ is the original signal at *nth* frame and *DCT* is discrete cosine transform, after applying pre-filtering and windowing.

At the first stage of feature extraction by MFC analysis, the input signal is broken down into overlapping frame blocks to comprehensively capture the signal temporal features and changes [9]. An unwanted side effect of the frame blocking process is that each frame will produce leakage effect (high frequency components at the end of each frame). To minimize the leakage effect and maintain continuity between frames, a windowing method with Hamming window is applied [8]. The size of each frame (l_w) depends on the sampling frequency (F_s):

$$l_w = 2^{0.03 \log_2 F_s} \tag{2}$$

After windowing, the Fourier analysis is performed on each frame, resulting in short-time Discrete Fourier Transform (DFT) [10]. Then the derived values are grouped together in critical bands and weighted by a series of triangular filter banks called mel-spaced filter bank, as shown in Figure. 1.

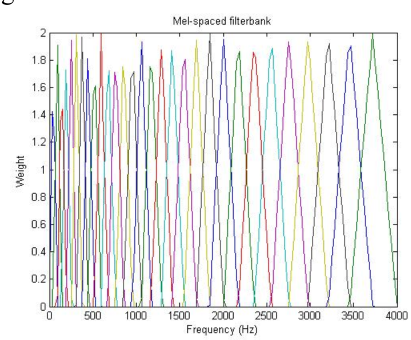

Figure 1. A 30 triangular mel-spaced filter banks.

The mel-spaced filter banks are designed based on the mel-scale frequencies, which mimic the human auditory system [8]. The human auditory system can detect frequency tones lower than 1 kHz in linear scale and frequencies higher than 1 kHz in logarithmic scale. Hence, the mel-scale frequencies are distributed linearly in the low frequency range but logarithmically in the high frequency range. The number of mel filter banks is adjusted according to the sampling frequency of the signals. The mel-scale frequency is given by:

$$mel = 2595\, log_{10}\left(1 + \frac{freq}{700}\right) \tag{3}$$

(3)

B. Binary Particle Swam Optimization

The Binary Particle Swarm Optimisation is a binarized modification of the original PSO algorithm. The original PSO algorithm is a stochastic optimization algorithm that iteratively searches for solutions in the problem space by taking advantage of the co-operative and competitive behavior of simple agents called particles. The algorithm search is directed by its velocity equation:

$$\begin{aligned} V_{id} = V_{id} &+ C_1(pBest - X_{id}) \times rand_1 \\ &+ C_2(gBest - X_{id}) \times rand_2 \end{aligned} \tag{4}$$

Subsequent to the new calculated velocities, X_{id} is updated according to Eq. (5).

$$X_{id} = X_{id} + V_{id} \tag{5}$$

where X_{id} and V_{id} are the particle position and velocity; *pBest* and *gBest* are the current best fitness and solution achieved by the swarm; C_1 and C_2 are the cognition and social learning rate; $rand_1$ and $rand_2$ are random number between '0' and '1'.

The PSO algorithm is modified to solve a binary problem, by representing the particle position as a probability of change instead of the actual solution [6]. Here, the feature selection problem is defined as:

$$\mathrm{F}_{\mathrm{reduced}} \subseteq \mathrm{F} \tag{6}$$

F consists of *n*-columns representing *n*-coefficient. To select a feature subset ($F_{reduced}$) using BPSO, a binary string of length (1 x n) is defined, so that each column has a bit assigned to it. The initial value of the binary string can be randomly defined during initialization. A value of '1' will be assigned to the binary string to show the column is to be considered in the construction of $F_{reduced}$. On the other hand, the value of '0' is assigned to show the column is ignored.

In the swarm, each particle carries a (1 x m) vector in solution, X_{id}. This vector contains the probabilities of change as mentioned earlier. During optimization, the vector elements will be used as a reference to alter the binary string from its initial state zero. If the particle vector element is more than 0.5, the binary bit will be changed to '1', otherwise the bit is maintained unchanged.

III. METHODOLOGY

The healthy and hypothyroid cry signals were acquired from the Instituto Technologico Superior de Atlixco dataset and the Chillanto dataset. The collection of cry signals was from infants of early born to 7 months. The files were recorded in .wav format. There were a total of 45 infant cry samples: 20 comes from the healthy infants while 25 comes from the hypothyroid infants.

A. Feature Extraction

In this study, the MFC coefficients were used to extract features from the cry signals. In order to obtain the MFC coefficients, the cry signals were first sampled at a sampling frequency of 8,000 Hz and then segmented at one second per segment. The segmented signals were then divided into frame of equal length, measured at approximately 16 ms per frame, with 50% overlap between them. The overlap was required in order to capture the temporal characteristic of the signals. To bridge the discontinuity between each frame, each block was multiplied with a Hamming window, with a width of 15.5 ms, which is the same as the length of each block.

Then, the Fast Fourier Transform (FFT) algorithm was applied to each frame to obtain the frequency spectrum.

These spectra were then filtered through a set of triangular bandpass filters known as the mel-scale filter banks to remove the linearity in each frame. Next, the logarithm of the amplitude spectrum (LAS) enclosed within the frequency scale was computed. The LAS describes the spectral envelope of the signal by showing how the spectral envelope varies with time over all the frames. The filter banks then executed the conversion from these frequency scales into mel-frequency scales. The number of filter banks defines the dimension of the MFC feature coefficients. Finally, the DCT algorithm was applied on LAS to reduce the number of parameters in the derived signal. By performing DCT of the logarithmic magnitude of the spectral energy, a set of MFC coefficients was obtained.

B. Feature Selection

To perform the BPSO, the number of filter banks and the number of MLP hidden units must be constant. In our work, only selection of coefficients was performed using BPSO. The optimum number of filter banks and hidden units used were 36 and 5 respectively, which were adopted from our previous work [11].

The swarm size of the BPSO was set to 100 to house a large solution space of 235. As the number of coefficients resulting from 36 filter banks was 35, the maximum iteration was fixed to 100. The initial value of the binary string was set to zero. Both C_1 and C_2 were set to offset the influence of the individual particle and the swarm in updating the position. The values of X_{max} and X_{min} were set to '1' and '0' respectively, so that the bit change probabilities would be bounded within this range. The values of V_{max} and V_{min} were set to '+1' and '-1' respectively, so that the maximum dynamic movement would be restricted within this range. The fitness function is a minimization function given by:

$$fitness = \frac{100 - CA_i}{n} \quad (7)$$

where CA_i is the classification accuracy of i-network and n is the number of neural networks tested which equals 50.

C. Classification of Hypothyroidal Infant Cry

In the classification of hypothyroidal infant cry, a three-layer MLP structure was used, with sigmoid transfer functions in both the hidden and output layers. The architecture consists of an input feature vector, 5 hidden nodes and 2 output neurons. The scaled conjugate gradient was used as the training algorithm. The MLP training process was repeated 50 times and the classification accuracy quoted was the mean. This technique was employed to minimize the effect from random weight initializations on the MLP classification performance.

IV. RESULTS & DISCUSSIONS

With reference to Figure 2, when the convergence of BPSO is experimented using 5 different seeds, different sets of results are produced. It is expected for all the trials to arrive at same result (fitness) at the end of the iterations. It is believed that the difference is due to the small number of particles used, relative to the size of solution space. It is also observed that the initial fitness values for all the cases are extraordinarily high. This is due to the initial random solutions. However, the initial fitness values gradually decline as iteration progresses, as better solutions are found.

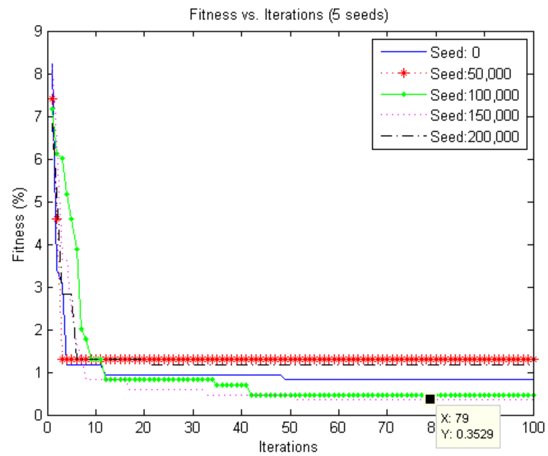

Figure 2. Fitness over iterations for 5 different seeds.

From Table 1, it can be observed that the best fitness is obtained with 150,000 seeds, after 32 iterations. It is fortunate that results from all the different seeds converge to a small fitness value within a close range, and hence are all acceptable. If a suitable number of particles and iteration can be found, an optimal result can be achieved.

TABLE I
FITNESS FOR 5 DIFFERENT SEEDS

No	Seed	Fitness
1	0	0.8235
2	50000	1.294
3	100000	0.4706
4	150000	0.3529
5	200000	1.176

With reference to Figure 3, by expanding the result from the best seed (150,000), it is found that the fitness function at the worse, average and best particle fitness, all converges to the lowest fitness values possible respectively. This shows that all particles try to outperform their individual best and global best as an improvement to their own solution.

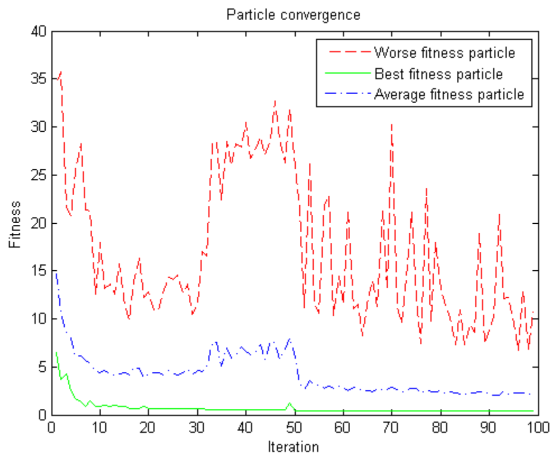

Figure 3. Convergence of particles with iterations for seed of 150,000.

Figure 4 shows the frequency of coefficients being selected by global best (*gBest*) in the case for seed of 150,000. Big differences can be observed in these frequencies amongst the different coefficients. It is found that the first, fourth, eleventh and fourteenth coefficient are coefficients popularly included, each time particles is updated as global best particles. This means that significant features are selected many fold more than the insignificant ones. This feature subset can then be used as the basis to significant features from MFC analysis. Hence, the significant MFC coefficients chosen for recognizing cries from infants with hypothyroidism are coefficients 1, 4, 7, 11, 14, 17, 18, 22, 29, 30 and 34. It is found that the classification accuracy produced with this selection of coefficients is 99.65%.

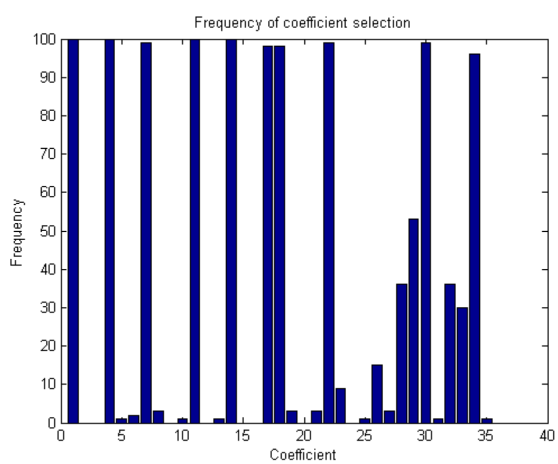

Figure 4. Frequency of coefficients selected by gBest for seed of 150,000.

V. CONCLUSIONS

The effectiveness of Binary Particle Swarm Optimisation as a feature selection technique for recognizing infant cries with hypothyroidism has been discussed in this paper. Our work has shown that the Binary Particle Swarm Optimisation algorithm is effective in enhancing the accuracy of the MLP classifier. This algorithm has succeeded as an automated heuristic to select significant coefficients for recognising cries from infants with hypothyroidism. A high classification accuracy of 99.65% is obtained for the MLP classifier with 36 filter banks, 5 hidden nodes and 11 BPS-optimised MFC coefficients.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Science, Technology and Innovation (eScience Grant 06-01-01-SF0210) and Universiti Teknologi Mara, Malaysia, for the facilities provided to carry out this research project. The authors would also like to thank Dr. Carlos A. Reyes-Garcia, Dr. Emilio Arch-Tirado, Dr. Edgar M. Garcia-Tamayo and the INR-Mexico group for their dedication in collecting the Baby Chillanto Database as well as Instituto Technologico Superior de Atlixco of Mexico for their database on infant cry signals. The Baby Chillanto Data Base is a property of the Instituto Nacional de Astrofisica Optica y Electronica – Conacyt, Mexico.

REFERENCES

[1] O. F. Reyes-Galaviz and C. A. R.-. Garcia, "A System for the Processing of Infant Cry to Recognize Pathologies in Recently Born Babies with Neural Networks," in 9th Conference Speech and Computer (SPECOM), St. Petersburg, Rusia, pp. 1-6, 2004.

[2] N. Setian, "Hypothyroidism in children: diagnosis and treatment," Jornal de Pediatria, vol. 83, pp. 209-216, 2007.

[3] D. Lederman, "Automatic Classification of Infants Cry," M.Sc. Thesis, Department of Electrical And Computer Engineering, Ben-Gurion University of The Negev, Negev, Israel, 2002.

[4] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, "A Comparison of Neural Network Architectures for the Classification of Three Types of Infant Cry Vocalizations," in Annual International Conference of the IEEE Engineering in Medicine and Biology, Montreal, pp. 821-822, 1995.

[5] M. Petroni, A. S. Malowany, C. C. Johnston, and B. J. Stevens, "Classification of Infant Cry Vocalizations Using Artificial Neural Networks," in International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95, pp. 3475-3478, 1995.

[6] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in Proc. 1997 Conf. Systems, Man, Cybernetics, Piscataway, NJ, pp. 4104–4108, 1997.

[7] I. M. Yassin, M. N. Taib, H. A. Hassan, A. Zabidi, and N. M. Tahir, "Heat Exchanger Modeling using NARX Model with Binary PSO-based Structure Selection Method," in International Conference on Computer Applications and Industrial Electronics (ICCAIE 2010), Kuala Lumpur, 2010.

[8] A. Zabidi, W. Mansor, L. Y. Khuan, R. Sahak, and F. Y. A. Rahman, "Mel-Frequency Cepstrum Coefficient Analysis of Infant Cry with Hypothyroidism " in 2009 5th Int. Colloquium on Signal Processing & Its Applications, Kuala Lumpur, Malaysia, pp. 204-208, 2009.

[9] J. L. Brock, "Acoustic Classification using Independent Component Analysis," M.S. Thesis, Department of Computer Science, Rochester Institute of Technology, Rochester, NY, 2006.

[10] D. S. Williamson, "Automatic Music Similarity Assessment and Recommendation," M.S. Thesis, Faculty of Electrical Engineering, Drexel University, Philadelphia, PN, 2007.

[11] A. Zabidi, W. Mansor, L. Y. Khuan, I. M. Yassin, and R. Sahak, "Investigation of Mel Frequency Cepstrum Coefficients parameters for classification of infant cries with hypothyroidism using MLP classifier," in The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1-4, 2010.