Identifying Pain and Hunger in Infant Cry with Classifiers Ensembles

Sandra E. Barajas-Montiel
Student Member
sandybarajas@ccc.inaoep.mx

Carlos A. Reyes-García
Member
kargaxxi@ccc.inaoep.mx

Abstract

The present work presents the experiments performed with two kinds of ensembles to classify infant cry. The ones selected for testing during the presented experiments are: a Boosting Ensemble of Artificial Neural Networks and a Boosting Ensemble of Support Vector Machines. The design and implementation of the ensembles as well as the experiments and some of the results are shown. The experiments are aimed to classify the types of pain - no pain and hunger - no hunger cries.

1. Introduction

Babies express their needs and feelings through cry. Several studies have demonstrated that infant cry is a powerful tool that can be used to decipher the needs of the babies or for making medical diagnoses of pathologies at very early stages of life.

The first works with infant cry were initiated by Wasz-Hockert since the beginnings of the 60's [1]. In 1964 the research group of Wasz-Hockert showed that the four basic types of cry can be identified by listening: pain, hunger, pleasure and birth [1].

Since then, many other studies related to this line of research have been reported. Sergio D. Cano has carried out and directed several works devoted to the extraction and automatic classification of acoustic characteristics of infant cry. In one of those studies, in 1999 Cano presented a work in which he demonstrates the utility of the Kohonen's Self-Organizing Maps in the classification of Infant Cries Units [2].

In Reference [3] a radial basis function (RBF) network is implemented for infant cry classification in order to find out relevant aspects concerned with the presence of CNS diseases. First, an intelligent searching algorithm combined with a fast non-linear classification procedure is implemented, establishing the cry parameters which better match the physiological status previously defined for the six control groups used as input data. Finally the optimal acoustic parameter set is chosen in order to implement a new non-linear classifier based on a radial basis function network, an ANN-based procedure which classifies the cry units into a 2 categories, normal-or abnormal case.

Reyes and Orozco [4] classified cry samples from deaf and normal babies with feed-forward neural networks, reporting a 97.43% of efficiency. In reference [5] it is showed the implementation of a Fuzzy Relational Neural Network for Detecting Pathologies by Infant Cry Recognition with percentage of correct classification of 97.3% and 98%. Petroni, Malowany, Johnston and Stevens [6] classify cry from normal babies to identify pain with different artificial neural networks. They reported 61% up to 86.2% of precision in the classification.

Among the models for pattern classification, neural networks have been used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques.

Another recently becoming popular classification technique is the Support Vector Machine (SVM), which was developed by Vapnik and his group at AT&T Bell laboratories [7]. SVM's have been used for isolated handwritten digit recognition, object recognition, speaker identification, charmed quark detection, face detection in images, and text categorization. Experimental results indicate that SVM's can achieve a generalization performance that is greater than or equal to other classifiers, while requiring significantly less training data to achieve good results [8].

An important and recent approach to improve performance is to form ensembles. The predictions of the classifiers members of the ensemble are averaged or combined by voting to form the ensemble's prediction.

Frequently, ensemble formation is combined with re-sampling of the data set. This approach can significantly improve generalization performance.This paper presents a Feed Forward Neural Network Boosting Ensemble and a Support Vector Machine Boosting Ensemble for the automatic classification of infant cry to identify pain and hunger.

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

2. The automatic infant cry recognition Process

The Automatic Infant Cry Recognition (AICR) process is very similar to the Speech Recognition Process. Basically the AICR can be divided in two sections. The first section is for signal processing and the second one is for pattern classification. In the signal processing phase, the cry signal is first normalized and cleaned, and then it is analyzed to extract the most important characteristics in function of time.

In AICR like in any pattern recognition problem, the goal is that given an input pattern we obtain as an output at the end of the recognition process the class to which this pattern belongs. Fig. 1 shows the general Automatic Infant Cry Recognition process.

2.1 Signal processing

The acoustical analysis of the raw cry wave form provides the information needed for its recognition. At the same time, it discards unwanted information such as background noise and channel distortion [9]. In this phase we make a transformation of measured data into pattern data. There are several techniques for analyzing cry wave signals, some of them are: Linear Prediction Coefficient (LPC), Mel Frequency Cepstral Coefficients (MFCC), Intensity, Spectral Analysis. For the described experiments we use Mel Frequency Cepstral Coefficients.

The MFCC are analog to perceptual characteristics. They can be obtained like filtered signals through different frequency scales. The Mel spectrum operates on the basis of selective weighing of the frequencies in the power spectrum. High order frequencies are weighed on a logarithmic scale whereas lower order frequencies are weighed on a linear scale. This technique pretends to simulate the characteristics that the ear has as a filter, which is more sensitive to some frequencies than to others [10], [11].

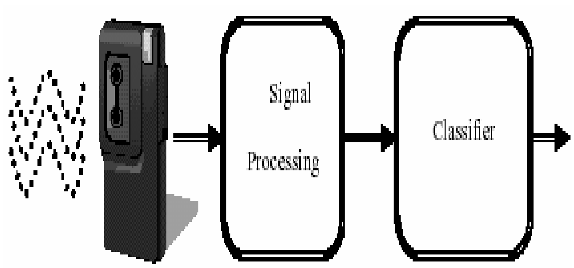

Fig.1. The automatic infant cry recognition process

The applications of this technique during signal processing results in the production of values of a set of acoustic features. The set of values for n features may be represented by a vector in an n-dimensional space. Each vector represents a pattern. Fig. 2 shows examples of MFCC extracted from hunger cries.

2.2 Pattern classification

In this phase we determine the class or category to which each cry pattern belongs to. The Infant Cry Recognition Process is a binary task, it is to say, the classifier must decide whether or not a cry belongs indeed to the class it claims to be.

For the present work development we'll focus on the description of the connectionist models, known as neural networks, and the description of support vector machines. We have selected this kind of models, in principle, because of their adaptation and learning capacity.

3. Support vector machines

Support vector machines (SVMs) have gained much attention since their introduction. This algorithm is an alternative training technique for Polynomial, Radial Basis Function and Multi-Layer Perceptron classifiers[12].

Support vector machines are based on the Structural Risk Minimization principle [13] from computational learning theory.

An SVM is a binary classifier that makes its decisions by constructing a linear decision boundary or hyperplane that optimally separates the two classes. It uses the discriminant function $f{:}X \subseteq \mathrm{R}^n$? R of the following form:

$$f(x) = < a \cdot k_S(x) > + b. \tag{1}$$

The $k_S(x) = [k(x, s_1), \ldots, k(x, s_d)]^T$ is the vector of evaluations of kernel functions centered at the support vectors $S = \{s_1, \ldots, s_d\}$, $s_i \in \mathrm{R}^n$ which are usually subsets of the training data. The $a \in R^l$ is a weight vector and $b \in R$ is a bias. The linear SVM aims to train the linear discriminant function:

$$f(x) = <w \cdot x> + b. \tag{2}$$

of the binary classifier [14]:

$$q(\boldsymbol{x}) = \begin{cases} 1 \text{ for } f(\boldsymbol{x}) = 0, \\ 2 \text{ for } f(\boldsymbol{x}) < 0. \end{cases} \tag{3}$$

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

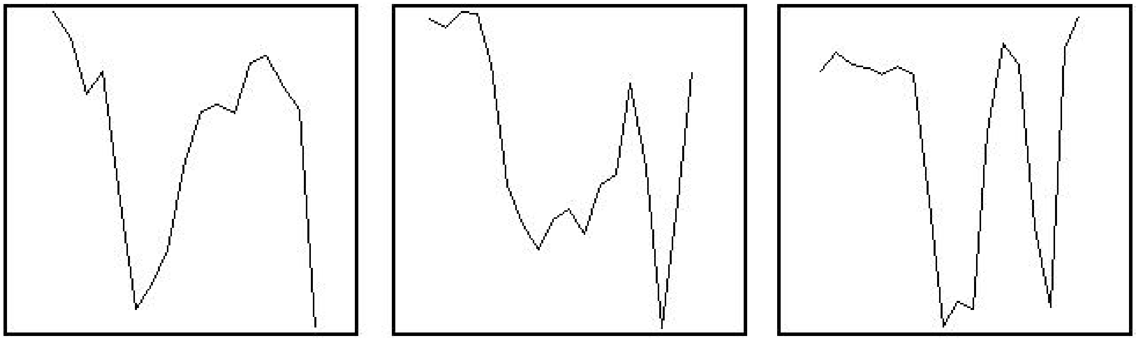

Fig. 2. Mel frequency cepstral coefficients extracted from hunger cries.

For linearly separable data labeled $\{x_i, y_i\}$, , $x_i \in R^d$, $y_i \in \{-1,1\}$, $i = 1\ldots N$ the optimal hyperplane is chosen according to the maximum margin criterion, i.e. by choosing the separating plane that maximizes the Euclidean distance to the nearest data points on each side of that plane.

This is achieved by minimizing the square of the L2-norm of w, $\|\mathbf{w}\|_2^2$ subject to the inequalities $(x_i \bullet w+b)y_i >= 1$ for all i.

The solution for the optimal hyperplane, w_0, is a linear combination of a small subset of data, $\mathbf{x}_s$, $s \in \{1 : : N\}_g$, known as the support vectors. These support vectors also satisfy the equality $((x_s \bullet w_{0+}b))y_s = 1$ [8].

4. Feed forward neural networks

Artificial Neural Networks are models based on the neural structure of the brain. The brain basically learns from experience and Neural Networks learn from experience too. These models have been widely used in the last decades. They can be applied across very different problems. Artificial Neural Networks are a successful solution to pattern recognition problems.

Basically, there are two reasons to use Artificial Neural Networks: They are capable to modelling complex functions and they can be applied to nonlinear problems. The other reason is, because they are easy to use.

Feed Forward Neural Networks are networks with a distinct layered structure, with all connections feeding forwards from inputs towards outputs. The typical network has an input layer, an output layer, and at least one hidden layer. Hornik KM, Stinchcombe M, White H. [15] showed that Feed Forward Artificial Neural Networks with one hidden layer can approximate any functions to any accuracy, but there is no theoretical limit on the number of hidden layers.

Each layer of the Neural Network is connected to the next layer; the input layer serves to introduce the values of the input variables, the hidden and output nodes calculate an activation value which is passed through an activation function to produce the output of each neuron. In Fig. 3 we can see the model of a typical Artificial Neural Network.

In general, the training can be supervised or not supervised. The methods of supervised training are those that are more commonly used, when labeled samples are available. Among the most popular models are the feed-forward neural networks, trained under supervision.

5. Boosting

Ensemble methods are aimed to improving the predictive performance of a given statistical learning or model fitting technique. We can say that an ensemble approach involves two tasks, the task for generating individual classifiers, and the task for combining individual predictions. Boosting algorithms were originally introduced by Freund and Schapire [16], [17], [18]. Boosting refers to a general and provably effective method of producing a very accurate prediction rule by combining rough and moderately inaccurate rules of thumb [19].

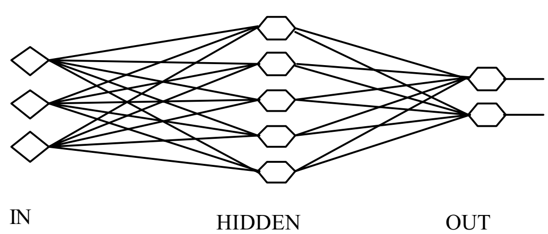

Fig.3. A typical artificial neural network.

The general idea of Boosting is to randomly generate training subsets, with proportional probability to the computed error from the original training set. The algorithm calls the base learning algorithm repeatedly using a different training subset, each time the base classifier generates a new prediction rule. Finally, the Boosting algorithm must combine these rules into a single prediction rule [20].

The boosting Algorithm AdaBoost was introduced in 1995 by Freud and Schapire [17], unlike the previous boosting algorithms of these two authors, AdaBoost needs no prior knowledge of the accuracies of the weak hypotheses. Rather, it adapts to these accuracies and generates a weighted majority hypothesis in which the weight of each weak hypothesis is a function of its accuracy; Fig. 4 shows pseudocode for AdaBoost in the generalized form given by Schapire and Singer [21].

6. Neural networks and support vector machines boosting ensembles

An artificial neural network ensemble is a learning paradigm where several artificial neural networks are used to solve a particular problem and their predictions are combined. The same approach is applied to the Support Vector Machine Ensemble. Some works [22] have shown that the generalization ability of an artificial neural work system can be significantly improved through an ensemble. Krogh and Vedelsby [23] conclude that the generalization ability of the applied ensemble is determined by the average generalization ability and the average ambiguity of the individual neural networks that constitute the ensemble.

In this paper we use the AdaBoost algorithm represented by means of the pseudocode showed in Fig. 4. The classifier called at each step of the ensemble is a feed forward neural network. Fig. 5 shows the general scheme for the Neural Network Boosting Ensemble used.

7. Implementation and results

For the present experiments we worked with samples of infant cries. The infant cries were collected by recordings done directly by medical doctors and then, each signal wave was divided in segments of 1 second, each segment represents a sample.

Then, acoustic features were obtained by means of Frequencies in the Mel scale (MFCC), by the use of the freeware program Praat v4.0.8 [24].

Every sample of 1 second is divided in frames of 50-milliseconds and from each frame we extract 16 coefficients, this procedure generates vectors whit 304 coefficients by sample. In order to reduce the dimensions of the sample vectors we apply Principal Component Analysis (PCA).

We used a corpus of 209 samples from pain cries, and 1418 samples representing the class no-pain, these samples include hunger, sleepy and uncomfortable types. Also we used a corpus of 759 samples of hunger cry, and 868 no-hunger samples.

The PCA algorithm, the Feed Forward Neural Network, and the AdaBoost algorithm were implemented in Matlab 6.5. For Support Vector Machines we use the Stprtool toolbox [14].

The parameter values for the Neural Networks Ensemble and the Support Vector Machines Ensemble were established after several heuristic experiments.

Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in X$, $y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$.
For $t = 1, \ldots, T$:

- Train base learner using distribution D_t.
- Get base classifier $h_t : X \to \mathbb{R}$.
- Choose $\alpha_t \in \mathbb{R}$.
- Update:

$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final classifier:

$$H(x) = \mathrm{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

Fig. 4. Pseudocode for the generalized version of the AdaBoost algorithm. From[20].

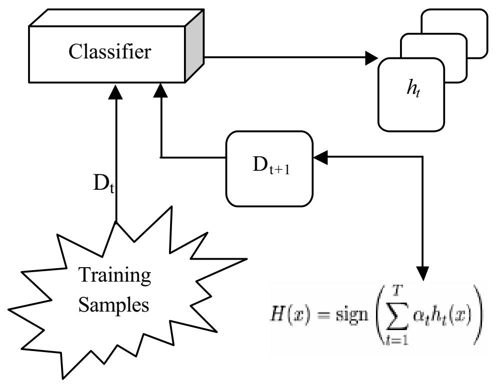

Fig.5. General scheme of the ensemble.

The number of hidden layers in the neural network is one with 3 hidden neurons; and the number of training epochs is 160. The Support Vector Machine uses the Radial Basis Function as kernel. The number of iterations for the ensemble is 5. For Neural Networks and Support Vector Machines we use ten fold cross validation.

7.1 Preliminary results

Two different classification experiments were performed, the first one consists in classify pain and no pain cry (P, N-P), the second one was made to classify infant cry in categories called hunger and no hunger (H, N-H). The results of these classification experiments using the Neural Network Ensemble are shown in Table I. In Table II we can observe the results obtained when the Support Vector Machine Ensemble was used. The classification accuracy was calculated by taking the number of samples correctly classified, divided by the total number of samples. In each classification task were used different number of principal components (PC's), the number of PC's tested in the experiments showed here was 2, 3, 10 and 16. The results are the average of five different experiments. When we use more Principal Components (PC's) we obtain better classification results, but when we use more than 16 PC's the correct classification percentage decreases.

With NN's ensemble we obtain 96.41% of efficiency when we in classify pain and no pain cry and with SVM's ensemble 85.16%. When we in classify hunger and no hunger with NN's ensemble we obtain 87.61% and with SVM's ensemble 76.41%. The variation of correct classification in each iteration of the SVM's ensemble was 2% or 3%, then the final results are no much different from the results when we use a single SVM. If we compare the results showed here with the results obtained in reference [6], which also identifies pain, we can see better performance with the use of ensembles of Neural Networks. With the reported experiments we confirm that it is easier to classify the tested cry patterns extracting MFCC features. Until now, we suppose that these results might be due to the fact that MFCC features are easier to differentiate.

8. Conclusions

Automatic Infant Cry Recognition is a powerful tool to determine the needs of the babies and their emotional status. An important approach to improve performance classification is to form ensembles. In this paper we present a NN and a SVM boosting ensemble for Automatic Infant Cry Recognition. The experiments show that best results were obtained with 16 PC's and with the Neural Network Ensemble. The different classification experiments showed here aim to represent more general problems than the problem to classify hunger and pain only. The most difficult to do is to collect more cry samples from normal babies, it is to say, babies without pathologies, to generalize better the classes no pain and no hunger. The correct classification we have obtained until this moment is very encouraging. Whit a larger number of samples we can generalize better our results in order to be closer to end up with a system that can be applicable to real life.

Table I.
Results of the classification using the neural network ensemble

Problem	Classification Accuracy			
	2 PCA	3 PCA	10 PCA	16 PCA
PN-P	88.27	92.10	94.01	96.41
HN-H	67.65	70.09	79.90	87.61

Table II.
Results of the classification using the support vector machine ensemble

Problem	Classification Accuracy			
	2 PCA	3 PCA	10 PCA	16 PCA
PN-P	85.16	85.40	83.97	85.16
HN-H	51.25	53.09	74.57	76.41

Proceedings of the 2005 International Conference on Computational Intelligence for Modelling, Control and Automation, and International Confe
Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05)
0-7695-2504-0/05 $20.00 © 2005 IEEE

9. Future work

In this work we use features extracted with MFCC technique, but other alternatives that can be useful are resonant frequencies, mean value, minimum and maximum of several signal features, or the combination of several acoustic features in the same pattern. On the basis of new approaches for speech recognition, other models or other types of ensembles also can be tested. We are in the process of testing new neural networks, and also testing new kinds of hybrid models, combining neural networks with genetic algorithms and fuzzy logic, or other complementary models. The collection of more samples will allow us to have a larger number of classes for the classification of normal cry and also for the classification of deafness levels. It is difficult to find the type of cry we require at the precise time, but with enough samples a classification of cry can be applied to detect pain, hunger, etc from deaf babies too.

10. Acknowledgments

This work is part of a project that is being financed by CONACYT-Mexico (C01-46753).

11. References

[1]Wasz-Hockert, O., Partanen, T., Vuorenkoski, V., Valanne, E. and Michelsson, K, "The Identification of Some Specific Meanings in Infant Vocalization," Experiencia, vol. 20, pp. 154-156, 1964.

[2]S.D. Cano-Ortiz, D. I. Escobedo-Becerro, "Clasificación de Unidades de Llanto Infantil Mediante el Mapa Auto-Organizado de Koheen," I Taller AIRENE sobre Reconocimiento de Patrones con Redes Neuronales, Universidad Católica del Norte, Chile,pp. 24-29, 1999.

[3] Sergio D. Cano Ortiz, Daniel I. Escobedo Beceiro, Taco Ekkel2, "A Radial Basis Function Network Oriented for Infant Cry Classification", Proc. Of 9th Iberoamerican Congress on Pattern Recognition, Puebla, Mexico, 2004.

[4]J., Reyes, C.A., Reyes, "Mel-frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry whit Feed-Forward Neural Networks," Proc. of ESANN, Bruges, Belgium, 2003.

[5]Suaste, I., Reyes, O.F., Diaz, A., Reyes, C.A., "Implementation of a Linguistic Fuzzy Relational Neural Network for Detecting Pathologies by Infant Cry Recognition," Proc. of IBERAMIA, Puebla, Mexico , pp. 953-962, 2004.

[6]M. Petroni, A.S. Malowany, C.C. Johnston, B.J. Stevens, "Identification of Pain from Infant Cry Vocalizations Using Artificial Neural Networks," Proc. of SPIE, vol. 2492, pp. 729-738, 1995.

[7]C. Cortes and v. Vapnik, "Support vector networks", Machine Learning, vol. 20, pp. 1-25, 1995.

[8]V. Wan, W.M. Campbell, "Support Vector Machines for Speaker Verification and Identification," IEEE International Workshop on Neural Networks for Signal Processing, Sydney, Australia, 2000.

[9]Livinson S.E., and Roe, D.B, "A Perspective on Speech Recognition," IEEE Communications Magazine, pp. 28-34, 1990.

[10]Orosco GJ, Reyes CA, "Mel-Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry with Feed-Forward Neural Networks," Proc. International Joint Conference on Neural Networks. Portland, Oregon, USA, pp. 3140-3145, 2003.

[11]Reyes O, Reyes CA., "Clasificación de Llanto de Bebés para Identificación de Hipoacusia y Asfixia por medio de Redes Neuronales," Proc. of the II Congreso Internacional de Informática y Computación de la ANIEL, Zacatecas, México, pp. 20-24, 2003.

[12]E. Osun, R. Freund, F. Girosi, Support Vector Machines: Training and Applications, Massachusetts Institute of Technology, Cambridge, MA, USA, 1997.

[13]Vladimir N. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York, 1995.

[14]Vojtech Fran, Vaclav Hlavac, Statistical Pattern Recognition Toolbox, Czech Technical University Prague, 1999.

[15]Hornik KM, Stinchcombe M, White H., "Multilayer Feedforward Networks are Universal Approximators," Neural Networks, vol. 2, pp. 359-366, 1989.

[16]Y. Freund, "Boosting a Weak Learning Algorithm by Majority," Information and Computation, vol. 121, pp. 256-285, 1995.

[17]Y. Freund, R.E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, pp. 119–139, 1995.

[18]R.E. Schapire, "The Strength of Weak Learnabylity," Machine Learning, vol. 5, pp. 197-227, 1995.

[19]Y. Freud, RE Schapire, "A Short Introduction to Boosting," Journal of Japanese Society for Artificial Intelligence, vol. 14, pp. 771-780, 1995.

[20]R.E. Schapire, "The Boosting Approach to Machine Learning An Overview," *MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[21]R.E. Schapire , Y. Singer, "Improved Boosting Algorithms Using Confidence-Rated Predictions," Machine Learning, vol. 37, pp. 297–336, 1999.

[22]Hansen LK, Liisberg L, Salamon P, "Ensemble Methods for Handwritten Digit Recognition," Proceedings of the IEEE-SP Workshop on Neural Networks for Signal Processing, pp. 333-342, 1999.

[23]Krogh A, Vedelsby J, "Neural Network Ensembles, Cross Validation, and Active Learning," Advances in Neural Information Processing Systems 7, Cambridge, MA: MIT Press, pp.231-238, 1995.

[24]Boersma, P., Weenink, D., Praat v 4.0.8: A System for Doing Phonetics by Computer. Institute of Phonetic Sciences of the University of Amsterdam, 2002.