Models and Analysis of Vocal
Emissions for Biomedical Applications
(MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

Neural Networks Techniques for Vocal Fold Pathology Detection.

Vyacheslav V. Parshin, Mikhail A. Toumilovich
Belarusian State University, Radiophysics Dept.
4, F. Skoryny Av., Minsk, 220050, Belarus
Tel: +375 172 771 007, Fax: +375 172 770 890,
E-mail: alxkrot@newman.basnet.minsk.by, mike@rfe.bsu.unibel.by

Keywords: speech processing, neural networks, training techniques, vocal fold pathology

The digital signal processing techniques are widely used in modern medicine and here there are many areas for investigations directed on enhancement of the effectiveness of these techniques and extension of their possibilities. One of such areas is the development of the automatic means for detection of human speech producing organs pathology based on analysis of their speech. Success in this area would supply the physicians with non-invasive procedure for speech pathology diagnostics that does not causes pain and discomfort to the patient and does not require the subjective evaluation. In this work the problem of speech pathology diagnosis is discussed and is referred as the speech signal classification problem. We suggest using the neural network approach to the problem because it is well developed for many similar problems in speech recognition [1,2].

As the input features for classification the short-time parameters of the speech signal can be used. Then the multilayer perceptron neural network classifier should be applied that allows to form the decision regions of any complexity. Due to the "dynamic" nature of the speech signal the useful information about the behavior of the acoustic parameters through time should be used to enhance the classification accuracy. This suggests to use more complicated "dynamic" networks such as time-delay neural networks or recurrent ones developed specially for time series classification [2]. The networks of this kind include time-delays or feedback links allowing the context-dependent processing of the speech.

As the training tool for the networks the error backpropagation procedure (EBP) or it generalizations are conventionally used. However it often converges too slowly and requires very large computational efforts for big size or "dynamic" neural networks. Moreover the procedure requires accurate tuning of its parameters in a heuristic manner and sometimes gets stick in a poor local minimum of the error function. It is expected that using alternative training techniques would allow to improve the performance of the EBP. It has been shown that the layer-wise training methods based on linearization of the neural net model and using least mean squares or stochastic identification methods often converge faster and do not include parameters that crucially govern the convergence properties [3]. This suggests to develop these methods for the training of the speech signal classifier neural networks.

Here we propose a stochastic training method for a neural network without feedback links or time delays based on the extended Kalman filter (EKF). The training is considered as the identification of the neural network model which is expressed by the following nonlinear system equations:

$$w(t+1) = w(t)$$
$$d(t) = h_t(w(t)) + v(t) = y(t) + v(t),$$

where the state vector $w(t)$ consists of all the linkweights of the network at time t, $d(t)$ and $y(t)$ are the desired and actual outputs at time t respectively, the error $v(t)$ is considered as white noise with covariance matrix $R(t)$, nonlinear function h_t is given by the structure of the neural network and its input vector. Using the EKF method of estimating the state vector given $\{h_s, R(s), d(s) \mid 0 \le s \le t\}$ leads to the following real-time learning algorithm:

$$\hat{w}(t) = \hat{w}(t-1) + K(t)[d(t) - \hat{y}(t)],$$
$$K(t) = P(t-1)H(t)^T[H(t)P(t-1)H(t)^T + R(t)]^{-1},$$
$$P(t) = P(t-1) - K(t)H(t)P(t-1),$$

where $K(t)$ is the gain matrix, $\hat{y}(t) = h_t(\hat{w}(t-1))$ - the estimate of $y(t)$ based on observations up to time t-1, $\hat{w}(t)$ - the estimate of the state vector based on observations up to time t, $P(t)$ expresses approximate error covariance, $H(t)$ - the matrix of the first derivatives of h_t.

The algorithm has too high computational complexity per pattern. However a simplified implementation of the EKF have been developed for the multilayer perceptron neural networks that allowed to reduce its computational complexity per iteration keeping the advantage of the fast convergence.

Computer simulations of the extended Kalman filter and error backpropagation training methods were performed for the problem of detection of the vocal fold pathology such as vocal fold nodules, dysphonia, chorditis and vocal fold paralysis. The database of sustained vowel sounds from healthy and vocal fold pathology patients was used. Spectral pathology component (SPC) was extracted from the short-time segments of the speech signals at the preprocessing stage using the EM-algorithm like that described in [4]. Mel-frequency spectral coefficients of the SPC were used as the input feature vector for the pathology detector. The multilayer perceptron neural network architecture was used in the simulations. Faster convergence of the EKF-based training techniques in comparison with the EBP was demonstrated.

References

1. Morgan N., Bourlard A. Proc. of the IEEE, May 1995, vol. 83, pp.742-769.
2. Lippman R.P. Neural Computation, 1989, vol.1, no.1, pp.1-38.
3. Rubanov N.S. Proc. IX European Signal Processing Conference, EUSIPCO-98, September, Greece, 1998.
4. Gavidia-Ceballos, J.H.L. Hansen. IEEE Transactions on Biomedical Engineering, vol. 43, No. 1, pp. 35-45, January 1996.