COMPARISON OF LINEAR AND NEURAL NETWORK-BASED POWER PREDICTION SCHEMES FOR MOBILE DS/CDMA SYSTEMS

X. M. Gao, J. M. A. Tanskanen, and S. J. Ovaska[1]

Laboratory of Signal Processing and Computer Technology
Helsinki University of Technology
Otakaari 5 A, FIN-02150 Espoo, Finland

[1] Laboratory of Telecommunications Technology
Helsinki University of Technology
Otakaari 5 A, FIN-02150 Espoo, Finland

Abstract—**This paper presents a novel neural network-based predictor for received power level prediction in direct sequence code division multiple access (DS/CDMA) systems. The predictor consists of a functional link neural network (FLNN) followed by a multilayer perceptron (MLP). An important but difficult problem in designing the cascade predictor is to determine the complexity of the networks. We solve this problem by using the predictive minimum description length (PMDL) principle to select the optimal number of input and hidden nodes. This results in a predictor with both good noise attenuation and excellent generalization capability.**

The optimized neural network predictor is compared with a class of FIR predictors with simulations employing noisy Rayleigh fading signals with 1.8 GHz carrier frequency. The results show that the neural predictor can provide a smoothed output signal with signal-to-noise ratio (SNR) gain of about 10 dB, which outperforms the linear counterpart. This makes the neural predictor well suitable for applications where 'delayless' noise attenuation and efficient reduction of fast fading are required.

I. INTRODUCTION

As the user capacity of a direct sequence code division multiple access (DS/CDMA) system is inherently interference limited, it is of paramount importance to keep the transmission power of each mobile user as low as possible [1]. This is crucial in the transmission from mobiles to a single base station, where all the mobile units need to be controlled by the base station. The feedback power control procedures allow the base station to send a power command to either lower or raise independently each user's transmitting signal power level, and so keep the received powers almost equal.

In a recent paper [2], a new power prediction scheme was presented for compensating the delays in the closed power control loop. In addition to delay compensation, the linear power predictor also bandlimits the noisy power signal and, therefore, both reduces its noise content and smoothens out fast fading. However, the noise attenuation capabilities of the computationally efficient linear schemes are very limited under low signal-to-noise ratio (SNR).

This paper presents a novel *hybrid* neural network-based predictor for efficient noise reduction. The neural predictor consists of a functional link neural network (FLNN) and a multilayer perceptron (MLP) [3]. An important but difficult problem in designing the cascade neural network is to determine the optimal structure needed for successful prediction and noise filtering. We use the predictive minimum description length (PMDL) principle to select the number of input and hidden nodes of the neural networks [4]. This results in a predictor with good noise attenuation, excellent generalization capability, and moderate computational complexity. The optimal predictor is then used with *on-line adaptation* for predictive filtering of the noisy power signal, the statistical characteristics of which may change with time.

The proposed neural predictor is compared with the linear prediction method of [2] with simulations. Both the neural predictor and the linear predictor use their optimized structures for power signal prediction. The results demonstrate that the neural predictor offers higher noise attenuation than the linear predictor. Besides, it is possible to design a neural predictor with considerably wider prediction bandwidth. The linear polynomial predictors with wide prediction bandwidth evidently have poor noise attenuation.

A brief overview of the applied channel model, the noise type, and the linear prediction schemes is given in Section II. The structure of the hybrid neural predictor and the PMDL principle for neural networks optimization are then introduced in Section III. In Section IV, the neural network-based predictor is applied to predictive filtering of a noisy power signal of a Rayleigh fading channel. Finally, we conclude this paper in Section V.

0-7803-3157-5/96 $5.00 © 1996 IEEE

II. NOISY FADING POWER SIGNAL AND LINEAR PREDICTIVE FILTERING SCHEMES

A. Channel Model and Noise

A detailed description of the modeling of a Rayleigh fading channel and noise was given by Jakes in [5]. In this paper, our simulator assumes the superposition of plane waves whose arrival angels are uniformly distributed. Different plane waves are associated with different Doppler shifts ranging from the minimum to the maximum specified by the mobile speed. The simulator consists of low frequency oscillators at these Doppler shift frequencies, and the frequency distribution results in a satisfactory approximation of the Rayleigh fading. The in-phase and quadrature components are formed by summing the appropriately weighted oscillator outputs. After multiplication with the corresponding carrier components, the signal is centered at the carrier frequency. Our carrier frequency was 1800 MHz, the sampling rate of the baseband equivalent in-phase and quardrature components was 1 kHz, and the applied vehicle speeds were 5 km/h and 50 km/h. To observe the smoothing properties of the predictors, the ideal reference signal used in calculating the SNR gain in the 50 km/h case was bandlimited to the frequency range of the 5 km/h case. The detailed implementation scheme was given in [2].

The noise used was zero mean white Gaussian noise that was independently added to the in-phase and quadrature components. In this paper, SNRs 10 dB and 0 dB within the components were chosen to be analyzed, to illustrate typical 'good' and 'bad' channels, respectively.

B. Linear Prediction Schemes for Noisy Fading Signal Prediction

The linear predictors often employed for predictive filtering of power signals are the *Heinonen-Neuvo* (H-N) predictors [6] and the recursive linear smoothed Newton (RLSN) predictors [7]. Due to their recursive nature, the RLSN predictors can offer much better noise attenuation than the H-N predictors with equal computational burden.

There exist two power prediction schemes: 1) direct prediction of the noisy power signal which has been calculated from the noisy components, and 2) computing the predictive estimates of the components separately and obtaining the power estimate by summing the squared values of these components. In most cases, the noise attenuation capability of the linear methods is not very efficient. The H-N linear predictor can obtain about 5 to 9 dB SNR gain at urban vehicle speeds.

III. NEURAL NETWORK-BASED PREDICTOR OPTIMIZATION AND PMDL PRINCIPLE

A. The Structure of the Neural Network-Based Predictor

Our neural network-based predictor is shown in Fig. 1. The predictor consists of two modules. An FLNN is used in the Module-1. The output of the FLNN is then fed to the input of the Module-2. An MLP with one hidden layer is used in the Module-2. The hyperbolic tangent sigmoid function is used as the nonlinear transfer function of the hidden nodes, and the transfer function of the output node is linear. The single node in the output layer represents the one-step-ahead prediction. A tapped delay line type input stage was used to make it possible to filter out the additive noise.

B. Optimization of the Structure of the Neural Networks Using the PMDL Principle

It is well known that a neural network with high complexity can maximize the mapping accuracy giving more precise output for the training data, but it may also give worse outputs for unseen data. On the other hand, networks with too few parameters cannot find out the mechanism that governs the data. Therefore, we must have a trade-off between the complexity and the generalization capability of the networks.

There exist several papers that discuss the selection of the number of nodes. For example, the Akaike Information Criterion (AIC) has been used to determine the number of input and hidden nodes [8]. However, the AIC has been proven inconsistent and has a tendency to overfit models. In this paper, we use the efficient PMDL method presented by Rissanen to optimize the neural networks of Fig. 1.

Rissanen proposed the shortest code description length for the model selection [9]. In subsequent papers [10], [11], this method gradually developed into stochastic complexity. We may regard the stochastic complexity as a generalization of Shannon's information or complexity. Indeed, for a given distribution $P(x)$,

$$I_s = -\log_2 P(x)\,, \tag{1}$$

evaluated at the observed data x, may be called Shannon's complexity of the data relative to a given distribution, or the model class $M = \{P(x)\}$.

For applications, the most important coding system is obtained from a class of parametric probability models:

$$M = \{f(x|\theta), \pi(\theta) \mid \theta \in \Omega^k, k = 1,2,\ldots\}, \tag{2}$$

where Ω^k is a subset of the k-dimensional Euclidean space with non-empty interior. Hence, there are k 'free' parameters. The stochastic complexity of x relative to the model class M is now given by:

$$\begin{aligned} I(x|M) &= -\log_2 f(x|M) \\ f(x|M) &= \int_{\theta\in\Omega^k} f(x|\theta)d\pi(\theta) \end{aligned} \tag{3}$$

Although the model class M includes the so-called 'prior' distribution π, its role is not the same as in the Bayesian inference. In fact, we need not select it at all, since we can construct a distribution $\pi(\theta|x)$ proportional to $f(x|\theta)$ [10]. The stochastic complexity represents the shortest code length attainable by the given model class. Frequently, for example in curve fitting and related problems, the models are not pri-

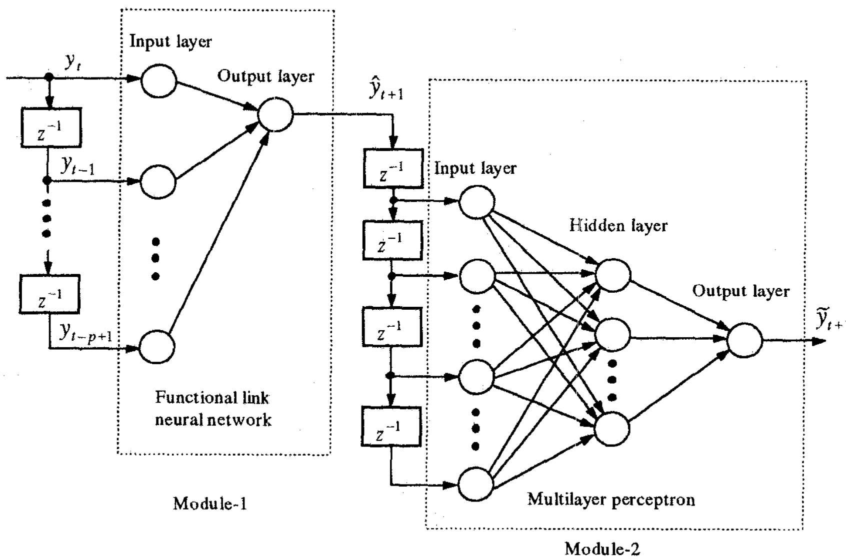

Fig. 1. The structure of the hybrid neural network-based predictor.

marily represented in terms of a distribution. Rather we can use a parametric predictor $\hat{y}_{t+1} = F(y|\theta)$ as in the case of neural networks, where $\hat{y}_{t+1}$ or $\tilde{y}_{t+1}$ is the output of the network, $y = (y_t, y_{t-1}, \ldots, y_{t-p+1})$ is the input, and θ denotes the array of all the weights as parameters.

In addition, there is a distance function $\delta(\varepsilon_{t+1})$ to measure the prediction error $\varepsilon_{t+1} = y_{t+1} - \hat{y}_{t+1}$ where y_{t+1} is the target output. Such a prediction model can be immediately reduced to a probabilistic model in which the optimization of ε_{t+1} causes the optimization of θ. In this case, we define the conditional Gaussian distribution

$$f(\hat{y}(t+1)|y(t),\theta,\sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-\varepsilon_{t+1}^2 / 2\sigma^2}. \quad (4)$$

The density (4) is then extended to a sequence as the code length

$$-\ln f(\hat{y}(t+1)|y^n,\theta,\sigma^2) = \frac{1}{2\sigma^2}\sum_{t=0}^{n-1}\varepsilon_{t+1}^2 + \frac{n}{2}\ln(2\pi\sigma^2), \quad (5)$$

where $y^n = (y_1, y_2, \ldots, y_n)$. After having fixed the model class, we have the problem to estimate the shortest code length obtainable with this class of models. Let $\hat{\theta}(y^t)$ and $\hat{\sigma}^2(y^t)$ be written briefly as $\hat{\theta}_t$ and $\hat{\sigma}_t^2$. They are the maximum likelihood estimates, i.e., the parameter values which minimize the code length $-\ln f(\hat{y}_{t+1}|y^t,\theta,\sigma^2)$ for the past data, in particular

$$\sigma_t^2 = \frac{1}{t}\sum_{t=1}^{t}\varepsilon_i^2. \quad (6)$$

Therefore, the predictive code length for the data is given by

$$-\ln(y^n|k) = \frac{1}{2}\sum_{t=0}^{n-1}\left[\frac{\varepsilon_{t+1}^2}{\hat{\sigma}_t^2} + 2\ln\hat{\sigma}_t\right] + \frac{n}{2}\ln(2\pi). \quad (7)$$

In the PMDL algorithm, the network parameters need not be encoded and they can be calculated from the past string by an algorithm. The model costs are added to the prediction errors, and the over-fitting and under-fitting are penalized automatically. A detailed description of the PMDL principle in neural network optimization can be found in [12].

IV. SIMULATION RESULTS

In our simulations, we first optimize the neural network structures using a linearly sweeping sinusoidal signal. The optimized neural predictor structure is then used in real-time prediction with on-line adaptation.

A. Off-Line Optimization of Neural Predictor Structures

Due to the time-varying characteristics of the power response of the Rayleigh fading channel, it is dangerous to optimize the predictor structure for a specific power signal. Therefore, an artificial sweeping sinusoidal signal is designed to optimize the network structures. The frequency of the sweeping signal varies from 0 to 100 Hz and the sampling rate is 1 kHz. The noise used is zero mean white Gaussian noise with different variances. Two cases, the SNRs of 10 dB and 0 dB are considered.

The optimization of the neural predictor is made in two steps. First, the Module-1 is optimized with the PMDL principles using the input samples $y = (y_t, y_{t-1}, \ldots, y_{t-p+1})$. After the optimal structure of the FLNN is determined, the optimization of the MLP in the Module-2 is made in the similar way using the output of the Module-1. The optimization procedures were performed under two conditions, i.e., with component SNRs of 0 dB and 10 dB. The optimized FLNN and MLP cascade form our optimal neural predictor.

B. Real-Time Prediction with On-Line Adaptation

The learning algorithm is crucial to our optimal predictor. Off-line learning methods calculate the weights of the networks without any reference to the time order of the training data and they must collect all the data before training. Thus, this kind of methods are time-consuming and not practical for our purpose. Furthermore, they cannot successfully handle a signal whose characteristics change in some possibly unknown manner which is the case in mobile applications.

As the fading signals are highly nonstationary, the learning must be adaptive. In our predictor, the adaptation of the FLNN in the Module-1 uses the Widrow-Hoff algorithm [13], and the MLP in the Module-2 uses an improved on-line backpropagation algorithm based on [14].

Figures 2 and 3 show the code lengths of the FLNN and the MLP with different number of input and hidden nodes under the SNR of 0 dB. It is easy to find that the optimal structure of the FLNN has 17 input nodes and the MLP has 6 input and 4 hidden nodes. Similarly, the optimal structure under 10 dB has the FLNN with 10 input nodes, and the MLP with 16 input and 2 hidden nodes.

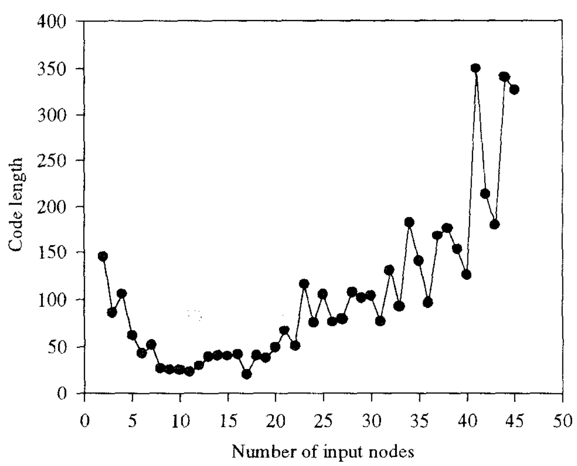

Fig. 2. Code lengths of different models of FLNN under component SNR = 0 dB.

The optimal predictor structure is then used for prediction of in-phase and quadrature components of the demodulated signal separately. The noisy, with the SNR of 0 dB, and the noiseless in-phase components are shown in Fig. 4, and the output of the optimal neural predictor is given in Fig. 5.

We use the measure

$$R_p = 10\log_{10}\left(\delta_s^2/\delta_p^2\right) \tag{8}$$

as the quantitative measure of the prediction performance. Here δ_s^2 is the mean squared value of the noiseless power signal, and δ_p^2 is the mean squared value of the prediction error of the predicted power signal. The optimal neural predictor can obtain about 11 dB SNR gain under the component SNR of 0 dB. The behavior under the component SNR 10 dB is similar, and about 8 dB improvement can be achieved. The results are 3 to 5 dB better than the maximum SNR gains of the H-N predictors. For the best results with the fixed-coefficient H-N predictors, the predictor length and the prediction scheme must be selected according to the vehicle speed and the current noise level. On the other hand, the adaptive neural predictor can maintain virtually the same SNR gain under varying conditions.

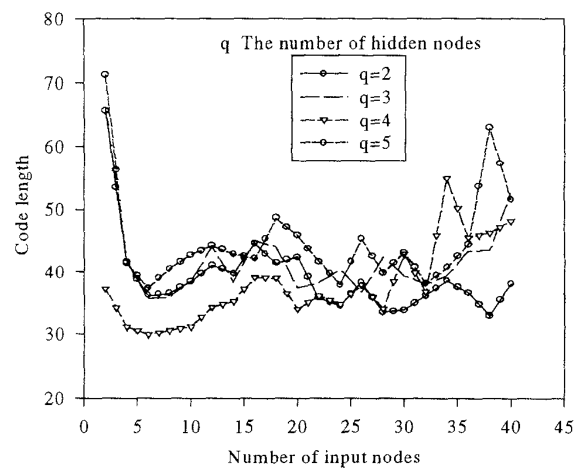

Fig. 3. Code lengths of different models of MLP under component SNR = 0 dB.

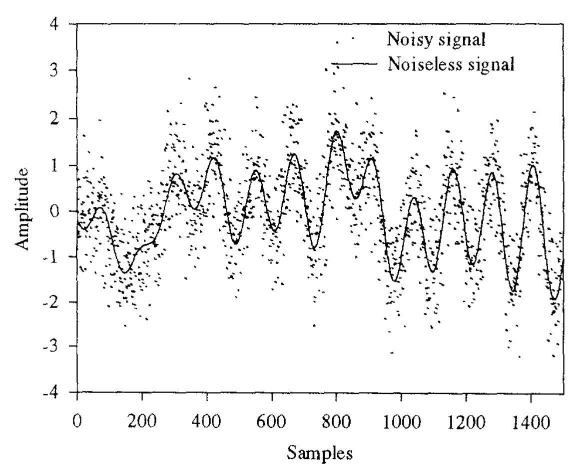

Fig. 4. Noisy and noiseless in-phase component of power signal at 5 km/h using Jakes' model.

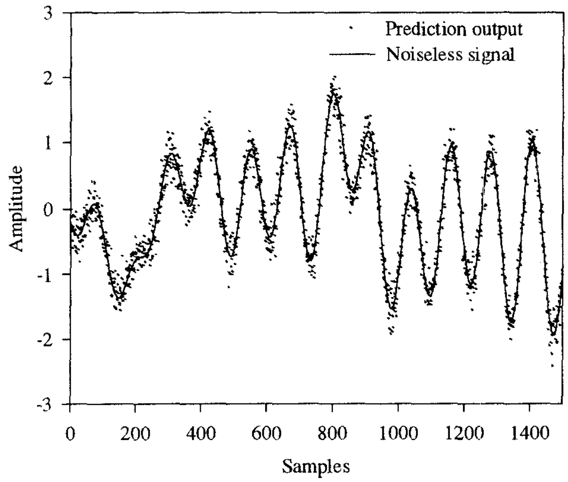

Fig. 5. The prediction of noisy in-phase component of power signal under 0 dB at 5 km/h.

In certain cases, the scheme 2) of predicting the in-phase and quadrature components separately may seriously increase the implementation costs. So, the scheme 1) would be an attractive alternative. The noisy power signal can be predicted directly by the optimal neural predictor. Now the SNR gains under the component SNRs of 0 dB and 10 dB are about 9 dB and 6 dB respectively, which are less than those of the scheme 2). Figure 6 shows the results of the power prediction under the component SNR of 0 dB.

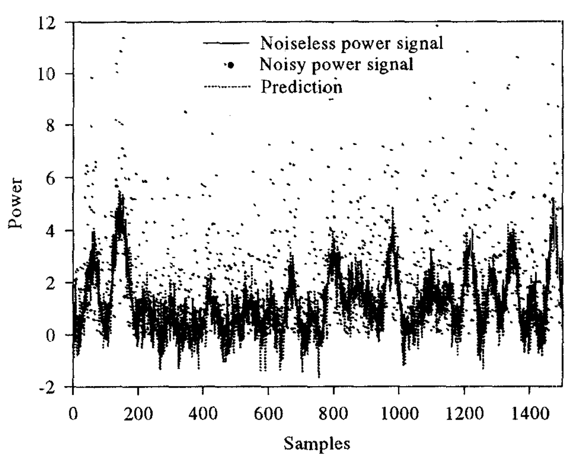

Fig. 6. Noisy and noiseless power signals, and the prediction of the optimal neural predictor under the component SNR of 0 dB.

The case of the mobile speed of 50 km/h is very similar to that of 5 km/h and the SNR improvements are comparable. However, the learning rate or adaptation time should be adjusted according to the mobile speed.

V. CONCLUSIONS

In this paper, we compared the linear and the neural predictors for received signal power prediction in DS/CDMA systems. Although the neural predictor has higher computational complexity than the linear approaches, it is still feasible from the application point of view, because the required sampling rate is only 1 kHz. In our experiment, the adaptation of the predictor requires 4250 floating point operations and prediction needs 1102 operations. The total computational requirement for predicting one sample is about 5000 arithmetic operations. Therefore, custom VLSI and DSP processors are the potential implementation platforms of our neural predictor. The presented neural predictor is also a natural pre-processing stage for advanced fuzzy and neural power controllers.

REFERENCES

[1] O. K. Tonguz and M. M. Wang, "Cellular CDMA networks impaired by Rayleigh fading: system performance with power control," *IEEE Trans. Vehicular Technology*, vol. 43, pp. 515-527, Aug. 1994.

[2] J. M. A. Tanskanen, A. Huang, T. I. Laakso, and S. J. Ovaska, "Prediction of received signal power in CDMA cellular systems," in *Proc. 45th IEEE Vehicular Technology Conference*, Chicago, IL, July 1995, pp. 922-926.

[3] D. E. Rumelhart and J. L. McCleelland (Eds.), *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, Vols. 1 & 2. Cambridge, MA: MIT Press, 1986.

[4] J. Rissanen, "Stochastic complexity and modeling," *Annals of Statistics*, vol. 14, pp. 1080-1100, 1986.

[5] W. C. Jakes (Ed.), *Microwave Mobile Communications*. New York: Wiley, 1974.

[6] P. Heinonen and Y. Neuvo, "FIR-median hybrid filters with predictive FIR substructures," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, pp. 829-899, June 1988.

[7] S. J. Ovaska and O. Vainio, "Recursive linear smoothed Newton predictors for polynomial extrapolation," *IEEE Trans. Instrum. Meas.*, vol. 41, pp. 510-516, Aug. 1992.

[8] D. B. Fogel, "An information criterion for optimal neural network selection," *IEEE Trans. Neural Networks*, vol. 2, pp. 490-497, Sept. 1991.

[9] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.

[10] J. Rissanen, "Universal coding, information, prediction, and estimation using predictive MDL principle," *IEEE Trans. Information Theory*, vol. 30, pp. 629-636, July 1984.

[11] J. Rissanen, "Stochastic complexity," *Journal of Royal Statistical Society, Series B*, vol. 49, pp. 223-239, 1987.

[12] M. Lehtokangas, "Modeling with layered feedforward neural networks", Doctoral Dissertation, Tampere University of Technology, Tampere, Finland, 1995.

[13] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[14] V. Ruiz de Angulo and C. Torras, "On-line learning with minimal degradation in feedforward networks," *IEEE Trans. Neural Networks*, vol. 6, pp. 657-668, May 1995.