Neural network modeling of ambulatory systolic blood pressure for hypertension diagnosis

Shahram Hosseini[1], Christian Jutten[2], and Sylvie Charbonnier[3]

[1] Laboratoire d'Acoustique, Metrologie, Instrumentation, 31062 Toulouse, France
hosseini@cict.fr
[2] Laboratoire des Images et des Signaux, 38031 Grenoble, France
Christian.Jutten@inpg.fr
[3] Laboratoire d'Automatique de Grenoble, 38402 St. Martin d'Hères, France
sylviech@lag.ensieg.inpg.fr

Abstract. A neural network model is used for estimating Ambulatory Systolic Blood Pressure (ASBP) variations from corporal acceleration and heart rate measurements. The temporal correlation of the estimation residual, modeled by a first order autoregressive (AR) process, is used for training the neural network in a maximum likelihood framework, which yields a better estimation performance. As data are collected at irregular time intervals, the first order AR model is modified for taking into account this irregularity. The results are compared by those of a neural network trained using an ordinary least square method.

1 Introduction

24-hours Ambulatory Systolic Blood Pressure (ASBP) recording has become a common tool to diagnose hypertension and evaluate anti-hypertensive drugs efficiency. The ASBP profile is obtained by a portable device which measures and records blood pressure during 24 hours, while the patient is leading a normal life. However, ASBP is a time varying variable which may increase significantly, in response to various parameters like emotional stress, physical activity, etc. This high variability makes the profile analysis and the evaluation of the drug treatment efficiency very difficult since the effect of the drug can be masked by sudden variations in ASBP due to external stimuli.

One of the solution proposed [1] for this problem is to equip the systolic blood pressure recorder with an accelerometer and an ECG portable recorder. They will provide body acceleration and heart rate data concomitant with ASBP measurements. Then, we will develop a model that can link ASBP variations with body acceleration and heart rate data. The estimation obtained by this model will be compared by the ASBP measurements, and used for reducing the variations due to changes in physical activities and, to a certain extent, to emotions.

In this paper, our objective is to estimate ASBP as a function of body acceleration and heart rate, using a neural network in a maximum likelihood framework.

As we will see, the estimation residual is temporally correlated. This correlation must be taken into account in maximum likelihood estimation. Since data are not collected in regular time intervals, the temporal correlation modeling requires some special consideration. Although the approach is developed for this special application, it is theoretically interesting and can be generalized to many other practical applications where the estimation residual is temporally correlated.

2 Experiments and data analysis

2.1 Description of the experiments

32 recordings of ASBP were achieved, from 16 different subjects[1]. For each subject, the experiment was repeated twice, separated by a period of a minimum of one week. Each recording lasted 24 hours, during which ASBP is monitored by a portable device. The ASBP is measured nearly every 15 minutes during the day and nearly every 30 minutes during the night. the intervals were not however completely regular and depend also on subjects. Body movement is measured continuously by a portable accelerometer (worn on the chest) which measures acceleration along the three axes. The acceleration signal is filtered between 0.05 and 10 Hz with two analog filters, and samples at a frequency of 50 Hz. The mean of the absolute value of the signal is measured continuously with a portable device.

2.2 Data analysis

For reducing the dynamic range, the logarithm of acceleration signal is computed. The heart rate over 3 minutes and the acceleration over 5 minutes preceding the ASBP measurements seem have the maximum correlation with ASBP. Thus, their averages over these periods are used in modeling. Data are also centered and divided by their standard deviations.

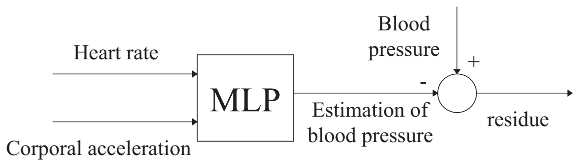
3 Neural network modeling

After the preprocessing of the previous section, for each subject, two data base corresponding to two experiments separated by at least one week are obtained. The first one is considered as the training data base and the second one as test data base. Some samples of training data base are used for forming a validation data set, used for choosing the network size. In the first modeling experiment, we use a two-inputs, one-output, multi layer perceptron (MLP) with one hidden layer. The output is the ASBP signal and the inputs are the normalized heart rate and body acceleration as explained in previous section.

---

[1] All the subjects were young students, male or female, between 20 and 30 years old, some suffering from mild hypertension.

To determine the optimal number of neurons in the network hidden layer, we could use a constructive algorithm [2]. However, since the constructive approaches may lead to sub-optimal solutions, we preferred to use a trial and error approach, especially as we have a validation data set. Thus, for each subject, the networks containing one to ten hidden neurons are trained by training data to minimize an Ordinary Least Square (OLS) criterion using a gradient descent algorithm. For each network, the validation error is evaluated using validation data and a cross-validation framework. The network providing minimum validation error is chosen as the optimal network. For one of the subject, the residual of this network estimation (as a function of time) and its autocorrelation function are presented in Fig. 1. The existence of the temporal correlation in the residual shows that the two above known inputs are not sufficient for removing the temporal correlation. It can be due to the influence of unknown temporally correlated sources (stress, feeding, static activities, ...). As we will show, this temporal correlation can be exploited by maximum likelihood method to improve the estimation.



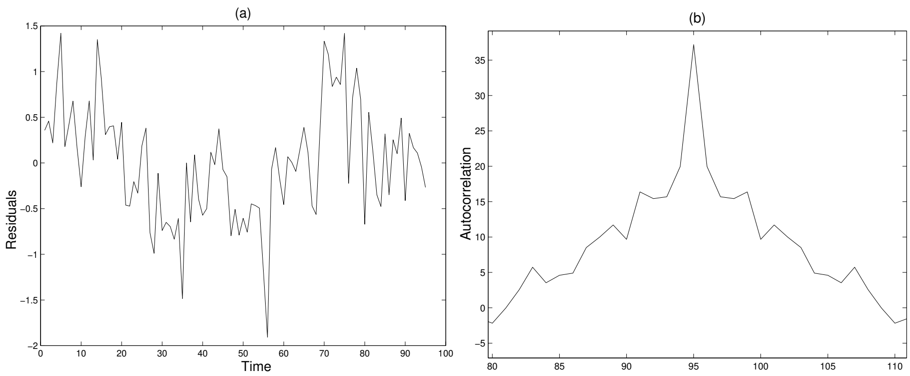


Fig. 1. Blood pressure modeling: Estimation residual and its autocorrelation function.

4 Maximum likelihood estimation

A typical non linear time series modeling using MLP can be presented as follows. Given a N-size data set, collected at time instances $t_1, t_2, \cdots, t_N$:

$$(\mathbf{x}(t_i), y(t_i)) = (\mathbf{x}(t_i), f(\mathbf{x}(t_i)) + \epsilon(t_i)) \ , \ i = 1, ..., N \tag{1}$$

where $\mathbf{x}(t_i)$ are samples of the input variable $\mathbf{x}(t) \in \mathcal{R}^P$, $y(t_i)$ are samples of the output variable $y(t) \in \mathcal{R}$ and $\epsilon(t_i)$ represent the samples of the output noise, one wants to find a good approximation of the function $f(.)$. The noise is defined as the part of y which does not depend on $\mathbf{x}$. Supposing $f(.)$ a continuous and bounded function, the universal approximation theorem of Cybenko [3] guarantees that $f(.)$ can be approximated with an arbitrary precision using a suitable size single hidden layer MLP whose output can be written as $g(\mathbf{x}(t), \mathbf{w}^*)$ where $\mathbf{w}^*$ is the optimal weight vector. In this case, (1) may be rewritten as:

$$(\mathbf{x}(t_i), y(t_i)) \simeq (\mathbf{x}(t_i), g(\mathbf{x}(t_i), \mathbf{w}^*) + \epsilon(t_i)) \ , \ i = 1, ..., N \tag{2}$$

Let $g(\mathbf{x}(t), \mathbf{w}^*)$ contains R sigmoidal hidden neurons and suppose $g(\mathbf{x}(t), \mathbf{w})$ represents the output of the class of single hidden layer MLP containing R sigmoidal hidden neurons; denoting $\epsilon(t, \mathbf{w}) = y(t) - g(\mathbf{x}(t), \mathbf{w})$, we would like to optimize $\mathbf{w}$ to obtain the best approximator for $\mathbf{w}^*$. When the noise samples, $\epsilon(t)$, are independent and identically distributed (i.i.d.) Gaussian variables, ordinary least square (OLS) estimation is usually used and coincides with asymptotically efficient maximum likelihood (ML) estimation. It consists in minimizing $E_{OLS} = \boldsymbol{\epsilon}(\mathbf{w})^T \boldsymbol{\epsilon}(\mathbf{w})$ where $\boldsymbol{\epsilon}$ is the vector containing all the samples of the residual. However, in many practical situations, like what we saw in ASBP modeling, the noise samples may be correlated in time. In this case, the OLS estimator is sub optimal and does not coincide with ML estimator . We have recently proposed [4] an algorithm for maximizing the likelihood function by taking into account the temporal correlation of the residual, which is modeled by an autoregressive process. This algorithm was however designed only for the case of regular data sampling, and can not be applied directly to our problem, because here the samples are collected at irregular time periods. In the following, for the special case of residual modeled by a first order autoregressive (AR1) process, we generalize the idea developed in [4] for irregular sampled data.

4.1 Theory

Inspired by the work of [5], the temporally correlated sequence $\epsilon(t)$ in the relation 1, will be described using the following non stationary AR1 model.

$$\epsilon(t_i) = \rho^{T_i} \epsilon(t_{i-1}) + n(t_i) \tag{3}$$

where $n(t)$ represents a sequence of Gaussian independent random samples and $T_i = t_i - t_{i-1}$. Note that if the sequence $\epsilon(t)$ is stationary: $\epsilon(t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$, the sequence $n(t)$ can not be stationary unless T_i is constant, because: $\sigma^2_{n(t_i)} =$

$(1-\rho^{2T_i})\sigma_\epsilon^2$, depends on T_i.

Our objective is to find the maximum likelihood estimation of the MLP weights $\mathbf{w}$, and the noise parameters ρ and σ_ϵ^2, which may be concatenated in a parameter vector $\boldsymbol{\theta} = [\mathbf{w}, \rho, \sigma_\epsilon^2]^T$. This estimation is obtained by maximizing the joint density of all the output observations:

$$L_1 = f_y(y(t_1), y(t_2), \cdots, y(t_N)) \tag{4}$$

Denoting the estimation residual by $\epsilon(t_i, \mathbf{W}) = y(t_i) - g(\mathbf{x}(t_i), \mathbf{w})$, where $g(\mathbf{x}(t_i), \mathbf{w})$ is the network output at instant t_i and $\mathbf{w}$ is the actual network weight vector, the above criterion can be rewritten as:

$$L_1 = f_\epsilon(\epsilon(t_1, \mathbf{w}), \epsilon(t_2, \mathbf{w}), \cdots, \epsilon(t_N, \mathbf{w})) \tag{5}$$

Considering (3), the above relation writes:

$$L_1 = f_\epsilon(\epsilon(t_1, \mathbf{w})) \prod_{i=2}^{N} f_\epsilon(\epsilon(t_i, \mathbf{w}) | \epsilon(t_{i-1}, \mathbf{w})) \tag{6}$$

or equivalently,

$$L_1 = f_\epsilon(\epsilon(t_1, \mathbf{w})) \prod_{i=2}^{N} f_n(\epsilon(t_i, \mathbf{w}) - \rho^{T_i} \epsilon(t_{i-1}, \mathbf{w})) \tag{7}$$

Since $\epsilon(t_1)$ and $n(t)$ are supposed Gaussian, we can write:

$$L_1 = \frac{\exp\{\frac{-\epsilon^2(t_1, \mathbf{w})}{2\sigma_\epsilon^2}\}}{\sqrt{2\pi\sigma_\epsilon^2}} \prod_{i=2}^{N} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2(1-\rho^{2T_i})}} \exp\{\frac{-[\epsilon(t_i, \mathbf{w}) - \rho^{T_i}\epsilon(t_{i-1}, \mathbf{w})]^2}{2\sigma_\epsilon^2(1-\rho^{2T_i})}\} \tag{8}$$

Taking the logarithm and after eliminating the constant terms, maximizing the likelihood function reduces finally to minimizing the following cost function:

$$L_2 = N \log \sigma_\epsilon^2 + \sum_{i=2}^{N} \log(1-\rho^{2T_i}) + \frac{1}{\sigma_\epsilon^2}[\epsilon^2(t_1, \mathbf{w}) + \sum_{i=2}^{N} \frac{[\epsilon(t_i, \mathbf{w}) - \rho^{T_i}\epsilon(t_{i-1}, \mathbf{w})]^2}{1-\rho^{2T_i}}] \tag{9}$$

It is sufficient then to replace $\epsilon(t_i, \mathbf{w})$ by $y(t_i) - g(x(t_i), \mathbf{w})$, and to minimize the cost function with respect to the parameter vector $\boldsymbol{\theta}$, using an iterative algorithm. We use the conjugate gradient method which finds the optimal value of the learning rate using a line minimization algorithm. To ensure that the estimation of ρ satisfies the condition $|\hat{\rho}| \leq 1$, we use the variable change $\rho = b/(1+|b|)$, and estimate b instead of ρ.
Note that if data are regularly collected, T_i is constant so that the cost function (9) may be considerably simplified.

5 Experimental results

For the 16 patients, the training data, preprocessed as explained in section 2, are used for training the one hidden layer MLP using both OLS and ML methods. For each patient and each method, the networks containing one to ten hidden neurons are trained, and the network providing minimum validation error is selected as the best network.

Figure 2 represents the variation of ASBP and its estimations by OLS and ML methods as functions of logarithm of corporal acceleration and heart rate for one of the patient. The extremely noisy nature of data can be remarked in this figure.

The training and the test mean square errors for the 16 patients are given in Table 1. The results confirm the better performance of ML estimator in generalization, thanks to the noise modeling. The OLS estimator pays no attention to noise structure and tries to memorize the noisy data. Hence, it has a better performance on the training data but a weaker generalization performance in comparison with ML estimator. Thus, the ML-trained networks are the better candidates for modeling ASBP.

Table 1. Training and test mean square errors for each of Ordinary Least Square and Maximum Likelihood estimators for the 16 subjects.

	training error		test error	
Patient	Least Square	Maximum Likelihood	Least Square	Maximum Likelihood
1	0.3581	0.3429	0.8141	0.7158
2	0.3076	0.3160	0.3818	0.3662
3	0.3838	0.3899	0.6675	0.5318
4	0.2514	0.3625	0.2364	0.2235
5	0.2550	0.2571	0.5917	0.5297
6	0.3185	0.3190	0.4953	0.4834
7	0.3261	0.3311	0.7162	0.7271
8	0.1583	0.1614	0.3903	0.4087
9	0.4555	0.5277	0.4745	0.4563
10	0.4205	0.4849	0.8057	0.7389
11	0.2790	0.2940	0.5317	0.5170
12	0.2075	0.2495	0.5425	0.5246
13	0.3073	0.3140	0.2684	0.2312
14	0.2631	0.2632	0.4832	0.4517
15	0.6797	0.7558	0.7194	0.6802
16	0.3680	0.4138	0.7339	0.6704
mean	0.3337	0.3614	0.5553	0.5160

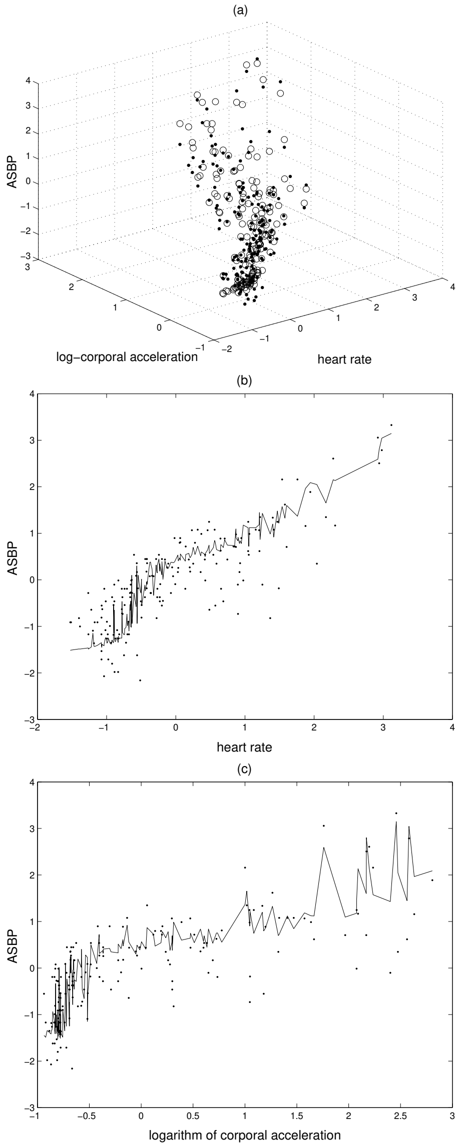


Fig. 2. Variation of blood pressure as a function of heart rate and logarithm of corporal acceleration, and its projections on two axes. (a),(b)et (c) OLS estimation. (d),(e) et (f) ML estimation. The points represent real data. The estimations are represented by the circles in the figures a and d, and by the solid lines in the other figures.

6 Conclusion

In this paper, we used a multilayer perceptron (MLP) for modeling ambulatory systolic blood pressure (ASBP) as a function of corporal acceleration and heart rate. We saw that whatever be the size of MLP, the estimation residual preserves some temporal correlation. It is because that the unknown temporally correlated factors influence ASBP. This effect was represented using an additive noise term, modeled by a first order autoregressive (AR1) model. To cope with the irregular data sampling, the AR1 model was modified to be non stationary. Afterwards, a maximum likelihood (ML) approach was used for estimating simultaneously the MLP weights and the noise model parameters. The experimental results showed that the ML estimator had a better generalization performance in comparison with the ordinary least square estimator. The approach, initially developed for this special application, may be used in many practical situations where the estimation residual is temporally correlated.

Acknowledgement The authors would like to thank the CHU of Grenoble for the medical data base used in this paper.

References

1. S. Charbonnier, S. Galichet, G. Mauris and J-P Siché, Statistical and fuzzy models of ambulatory systolic blood pressure for hypertension diagnosis. *IEEE Trans. on Instrumentation and Measurement*, vol. 49, no. 5,pp. 998-1003, Oct. 2000.
2. T.Y. Kwok and D.Y. Yeung, Constructive Algorithms for Structure Learning in Feedforward Neural Networks for Regression Problems. *IEEE Trans. on Neural Networks*, vol. 8, no. 3, May 1997.
3. G. Cybenko, Approximation by superposition of sigmoidal functions, *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303-314, 1989.
4. S. Hosseini and C. Jutten, Maximum likelihood neural approximation in presence of additive colored noise, *IEEE Trans. on Neural Networks*, vol. 13, no. 1, pp. 117-131, Jan. 2002.
5. R. F. Galbraith and J. I. Galbraith, On the inverse of some patterned matrices arising in the theory of stationary time series, *Journal of Applied Probability*, vol. 11, pp. 63-71, 1974.