Neural Nets WIRN11
B. Apolloni et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-972-1-178

A Collaborative Approach to Time-Series Prediction

Michele SCARPINITI [1], Danilo COMMINIELLO, Raffaele PARISI and Aurelio UNCINI

DIET - Department of Information Engineering, Electronics and Telecommunications, "Sapienza" University of Rome, via Eudossiana 18, 00184 Roma

Abstract. In this paper a collaborative filter combination for time-series prediction is considered. The basic idea is based on a convex combination of two kernel adaptive filters with different parameters. While the convergence of one filter is fast but not accurate, the convergence of the second one is much more accurate, even if slower. The convex combination of both filters allows to reach good performances in terms of convergence and speed. Some experimental results on the prediction of the Mackey-Glass time-series demonstrate the effectiveness of the proposed approach.

Keywords. Time-series prediction, Kernel adaptive filters, Collaborative filtering, Convex combination.

Introduction

Time-series prediction is a predominant application in many scientific branch, such as engineering, statistics, environmental science and so on. In literature several solutions to this problem are proposed: autoregressive (AR) models, autoregressive moving average (ARMA) models, artificial neural networks (ANN), etc. When time series has a stochastic nature, nonlinear models usually work better in the prediction task with respect to the linear ones [1,10]. ANNs are fully adopted in this way since they can be applied to complicated models being more flexible and universal, and having no special requirement for explicit underlying models.

Recently kernel adaptive filters (KAFs) have been introduced [11]. This solution is able to simply implement a nonlinear filtering. KAFs can be used in a wide range of applications and it was shown that they entail good performances in chaotic time-series prediction [11,14].

Some years ago a convex combination filter approach was proposed [2,3,4]. This approach can simply solve the ambiguity on the choice of the learning rate. In fact, the whole adaptive filter results in a combination of two single adaptive filters, the first with a huge learning rate and the latter with a small one. This solution overcomes the dichotomy between speed and convergence: while the second filter provides good convergence performances, the first one speeds up the convergence of the overall filter [3]. More recently

[1] Corresponding Author: Tel. +39 0644585869; e-mail: michele.scarpiniti@uniroma1.it.

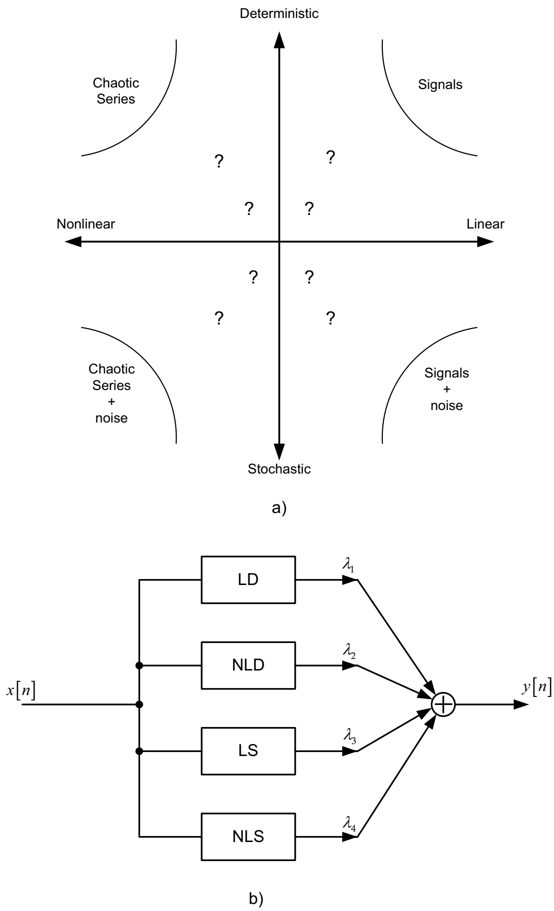

Figure 1. Possible variety of signals spanned by a certain degree of nonlinearity and uncertainty a) and respective filter combination for automatic selection of the relative subsystem b)

such a combination was successfully applied to several applications, such as nonlinear acoustic echo cancellation [5,7,6] and other signal processing problems.

The idea of filters combination is very interesting because it is possible to model a wide range of applications [13]. In fact real-world processes comprise both linear and nonlinear components, together with deterministic (that can be precisely described by a set of equations) and stochastic ones. In this way, models used to describe these real-world processes can be classified with a certain degree of nonlinearity and uncertainty, and described in a diagram (see Figure 1.a)). In literature only few cases as the linear stochastic ARMA and chaotic models are well understood, while real-world processes are often a combination of the previous four possibilities. In order to automatically take into account all the previous possibilities, a possible solution is to think to a system that automatically selects the right subsystem working on the relative quadrant (see Figure 1.b)). In this preliminary paper, we focus our attention on the use of a couple of kernel adaptive filters, in order to exploit the benefits of both the KAF and the convex combination. The adaptation of each kernel filter is performed by the classical LMS algorithm, known in this contest as Kernel LMS (KLMS) [11]. The consequent system is then adopted in the chaotic time-series prediction problem. Something similar was proposed in [16] adopting RBF networks. The experimental results are addressed to the short-time prediction of the Mackey-Glass (MG) chaotic time-series.

The paper is organized as follows: Section 1 introduces the concept of kernel adaptive filtering. Section 2 describes the convex combination of two adaptive filters while Section 3 shows some experimental results. Finally Section 4 draws our conclusions.

1. Kernel Adaptive Filters

Kernel adaptive filters (KAFs) are nonlinear filters whose parameters can be adapted during the learning process [11]. Kernel adaptive filters are implemented as standard (linear) adaptive filters where input data are first transformed into a high-dimensional feature space via a reproducing kernel. This transformation is performed directly by kernel evaluations. The mapping into a feature space can simplify the learning task.

The main issue in this approach is how to construct this mapping, in order to take into consideration the computational complexity and the generalization performances. This computation is evaluated through the use of kernels $K(\mathbf{x}, \mathbf{x}')$, where $\mathbf{x} = [x[n-N], \ldots, x[n-2], x[n-1]]^T$, N is the time delay length and $\mathbf{x}'$ is a new input. One of the most used kernel is the Gaussian one, defined as

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma \left\|\mathbf{x} - \mathbf{x}'\right\|^2\right), \tag{1}$$

where γ is a kernel parameter that can control the smoothing ability of the kernels. The Gaussian kernel has the universal approximating capability, is numerical stable and usually gives reasonable results [11].

The output of the i-th KAF filter is

$$y_i[n] = \mu_i \sum_{j=1}^{N} e_i[j] K\left(\mathbf{x}[j], \mathbf{x}[n]\right), \tag{2}$$

where μ_i is the learning rate and $e_i[n] = d[n] - y_i[n]$ is the error signal. The adaptation of each kernel filter is performed by the classical LMS algorithm, known in this contest as KLMS [11, chapter 2]. The KLMS algorithm update is computed without using the concept of weight, like for the classical LMS algorithm [9]. Instead it is present the sum of all past errors multiplied by the kernel evaluation of available data, as shown in eq. (2). For the choice of the learning rate μ_i, can be used the same criteria applied by adaptive filtering literature [9], since it has the role of a compromise between convergence speed and misadjustment.

The idea of the KAF is shown in Figure 2. This topology reminds us of a radial-basis function (RBF) network [8], except that the output weights are the prediction errors and $K(\cdot,\cdot)$ can be any Mercer kernel. In [11] it is also proven that KLMS algorithm shows the well-required self-regularization property, so it does not need regularization.

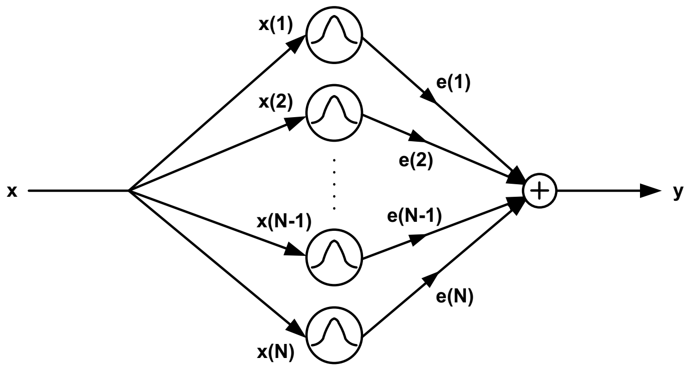

Figure 2. Network topology of the kernel adaptive filter (KAF)

2. Convex Combination of Kernel Adaptive Filters

Recently some interesting works have shown that some advantages can be taken from convex combination of two or more adaptive filters, working with different parameters and/or different learning algorithms [2,3,4]. This convex combination outperforms the single algorithms, improving the robustness of the overall architecture. In addition, it is well known that the combination of filters of different families of algorithms can improve the tracking capabilities of the whole system [15].

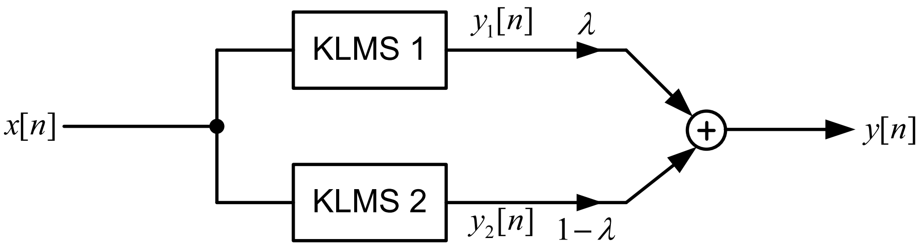

Figure 3. Convex combination of two adaptive KLMS filters.

As shown in Figure 3, the system output is evaluated as

$$y[n] = \lambda[n]y_1[n] + (1 - \lambda[n])y_2[n], \quad (3)$$

where $y_i[n]$ is the output of the i-th adaptive filter and $0 \leq \lambda[n] \leq 1$ is a parameter that selects the amount of the i-th output to the whole system output.

The adaptation of each kernel filter is then performed by eq. (2).

In order to reduce the gradient noise and to keep the mixing parameter in the range $[0, 1]$, the adaptation of $\lambda[n]$ can be carried out through the adaptation of another parameter, $a[n]$, related to $\lambda[n]$ by the following equation:

$$\lambda[n] = \left(1 + e^{-a[n]}\right)^{-1}. \quad (4)$$

The update of $a[n]$ is the given by:

$$a[n+1] = a[n] + \frac{\mu_a}{p[n]}\lambda[n](1-\lambda[n])e[n]\varepsilon[n], \qquad (5)$$

where $\varepsilon[n] = y_1[n] - y_2[n]$; the term $p[n] = \beta p[n-1] + (1-\beta)\varepsilon^2[n]$ is the estimated power of $\varepsilon[n]$ and β is a threshold close to one [5].

3. Results

We have tested our architecture for the short-time prediction of the Mackey-Glass (MG) chaotic time-series [12]. This time-series is derived from the following differential equation:

$$\frac{dx(t)}{dt} = -\beta x(t) + \frac{\alpha x(t-\tau)}{1 + x(t-\tau)^n}, \qquad (6)$$

with $\beta = 0.1$, $\alpha = 0.2$, $n = 10$ and $\tau = 30$. The eq. (6) was discretized at a sampling period of 6 seconds and a segments of 5000 samples was generated. Then a segment of 500 samples is randomly selected from the series and used as training data, while another 100 samples are used as test data. The data can also be corrupted by additive Gaussian noise with zero mean and standard deviation σ. As convergence criterion it is used the *Mean Square Error* or MSE versus the number of iterations n, defined by

$$\text{MSE} = E\left\{e^2[n]\right\}. \qquad (7)$$

For each training sample, the weights adaptation in eq. (2) is calculated and then the MSE is evaluated on the test set for this weight value. Hence three experimental tests are here proposed.

A first test is conducted on training data using a KLMS algorithm with parameters $\mu_1 = 0.2$, $\mu_a = 0.2$, $\gamma_1 = 1$ while $p[0] = 0.4$ and $\beta = 0.9$, in conjunction with a classical linear LMS algorithm with parameter $\mu_2 = 0.2$. The data are also corrupted by additive Gaussian noise with zero mean and standard deviation $\sigma = 0.01$. The MSE for this first experimental test is reported in Figure 4 while the predicted data of the test set is depicted in Figure 5. As we can see, the convergence of the whole output system is faster than the single ones, while the MSE tends to settle to the lowest one. Figure 4 clearly shows that the convergence is reached in about 75 iterations. Moreover Figure 5 shows that predicted samples are very close to the original ones.

A second test is conducted on training data using two KLMS algorithms with the same parameters $\mu_1 = 0.2$, $\mu_2 = 0.2$, $\mu_a = 0.2$, $\gamma_1 = 1$ and $\gamma_2 = 4$, while $p[0] = 0.4$ and $\beta = 0.9$. The data are also corrupted by additive Gaussian noise with zero mean and standard deviation $\sigma = 0.04$. The MSE for the second experimental test is reported in Figure 6, while the prediction error for the test set is depicted in Figure 7. As we can see from Figure 6, it is evident that also in this case the convergence of the whole output system is faster than the single ones, while the MSE tends to settle to the lowest one. Due to the higher value of additive noise, predicted samples move little from the original ones, as shown in Figure 7.

Finally a third test is conducted using a combination of two time series of MG type, with $\tau = 30$ and $\tau = 17$ (see eq. (6)), respectively. Two segments of 250 samples

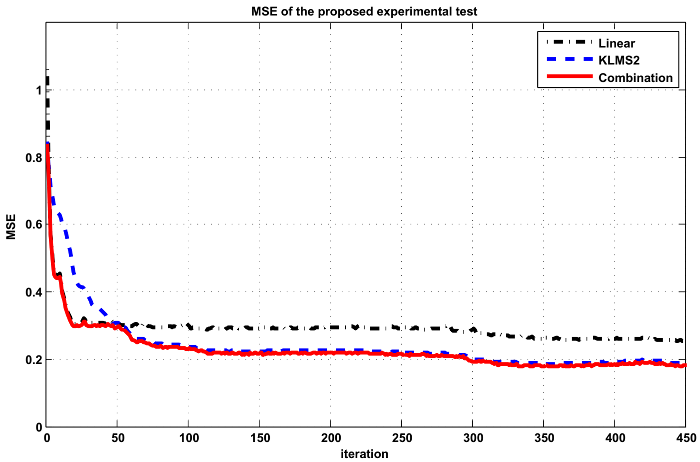

Figure 4. Mean Square Error (MSE) for the first experimental test

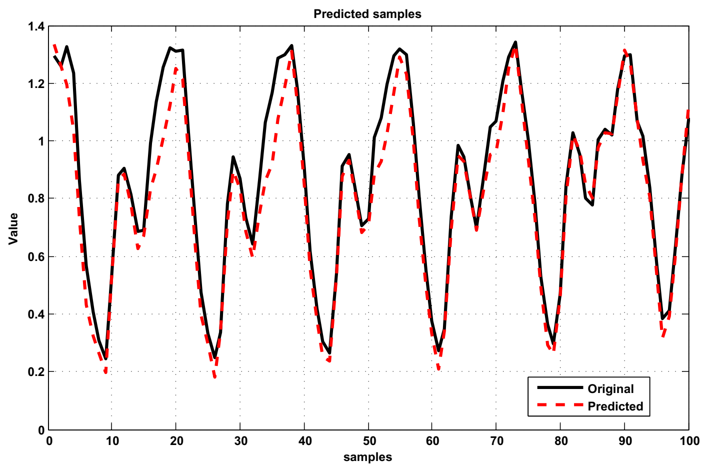

Figure 5. Predicted data of the test set in first experimental test

are randomly selected from each time-series and then combined in a training vector. Similarly it is done for the test set. The parameters are the same as the previous example, except $\mu_2 = 0.01$ and $\gamma_2 = 8$. The MSE for this experimental test is reported in Figure 8. In this third case, the MSE of the proposed system is following the lower one, even near and after the discontinuity due to the time-series changes. Also this experimental test confirms the effectiveness of the proposed approach.

4. Conclusions

In this paper we have introduced a novel approach based on a convex combination of kernel adaptive filters, applied to chaotic time-series prediction. The proposed approach can take advantage from both the simplicity and regularizing properties of the KAF and

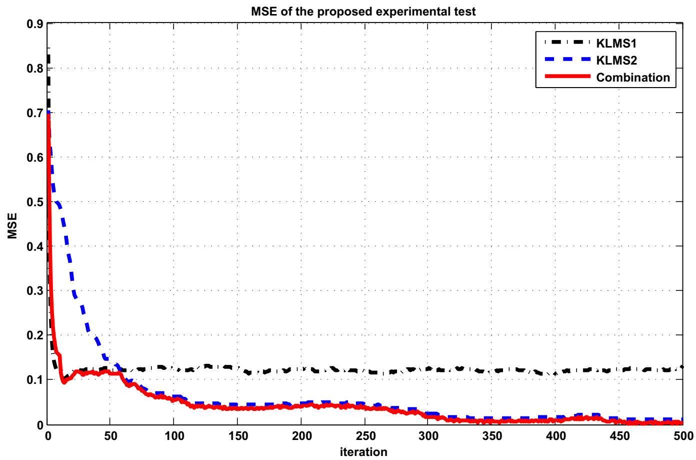

Figure 6. Mean Square Error (MSE) for the second experimental test

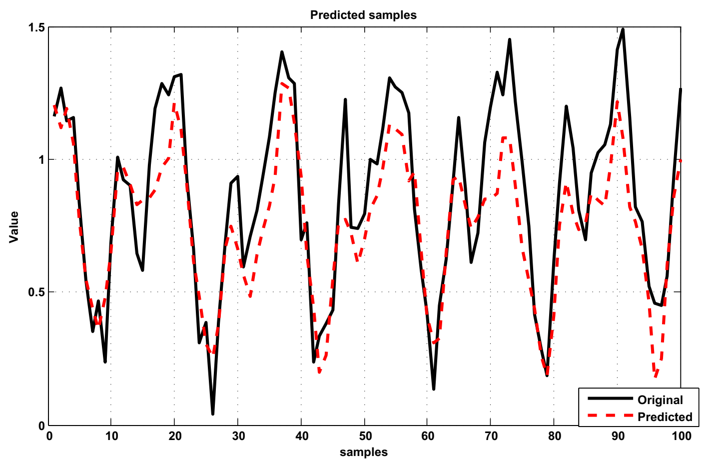

Figure 7. Predicted data of the test set in second experimental test

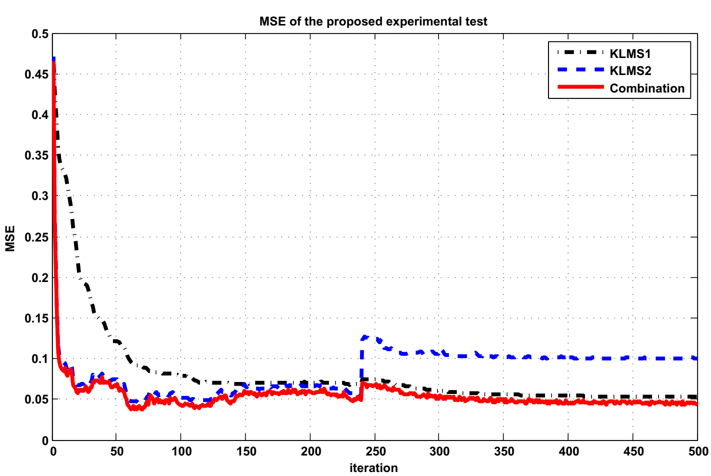

Figure 8. Mean Square Error (MSE) for the third experimental test

the robustness and fast convergence properties of convex combination of adaptive filters. The followed approach is able to successfully predict the values of the well known Mackey-Glass chaotic time-series, measuring the performance in terms of MSE. Different experimental tests were proposed in order to validate the proposed idea.

References

[1] H.D.I. Abarbanel, T.W. Frison, and L.S. Tsimring. Obtaining order in a world of chaos. *IEEE Signal Processing Magazine*, 15(3):49–65, May 1998.

[2] J. Arenas-Garcia and A. R. Figueiras-Vidal. Adaptive combination of normalised filters for robust system identification. *Electronics Letters*, 41(15):874–875, 2005.

[3] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed. Mean-square performance of a convex combination of two adaptive filters. *IEEE Transactions on Signal Processing*, 54(3):1078–1090, 2006.

[4] J. Arenas-Garcia, V. Gomez-Verdejo, and A. R. Figueiras-Vidal. New algorithms for improved adaptive convex combination of lms transversal filters. *IEEE Transactions on Instrumentation and Measurement*, 54(6):2239–2249, 2005.

[5] L. A. Azpicueta-Ruiz, M. Zeller, J. Arenas-Garcia, and W. Kellermann. Novel schemes for nonlinear acoustic echo cancellation based on filter combinations. In *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2009)*, pages 193–196, Taipei, Taiwan, April 19-24 2009.

[6] D. Comminiello, L. A. Azpicueta-Ruiz, M. Scarpiniti, A. Uncini, and J. Arenas-Garcia. Functional link based architectures for nonlinear acoustic echo cancellation. In *Proc. of Hands-free Speech Communication and Microphone Arrays*, Edinburgh, UK, May 30 - June 01 2011.

[7] D. Comminiello, M. Scarpiniti, R. Parisi, and A. Uncini. A functional link based nonlinear echo canceller exploiting sparsity. In *Proc. of International Workshop on Acoustic Echo and Noise Control (IWAENC2010)*, Tel Aviv, Israel, August 30 - September 2 2010.

[8] S. Haykin. *Neural Networks, a comprehensive foundation*. Prentice-Hall, 2nd edition, 1999.

[9] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, 4th edition, 2002.

[10] S. Haykin and J. C. Principe. Making sense of a complex world. *IEEE Signal Processing Magazine*, 15(3):66–81, May 1998.

[11] W. Liu, J. C. Principe, and S. Haykin. *Kernel Adaptive Filtering: A Comprehensive Introduction*. Wiley, 2010.

[12] M. C. Mackey and L. Glass. Pathological physiological conditions resulting from instabilities in physiological control system. *Science*, 197:287–289, 1977.

[13] D. P. Mandic, M. Chen, T. Gautama, M. M. Van Hulle, and A. Constantinides. On the characterization of the deterministic/stochastic and linear/nonlinear nature of time series. *Proceedings of the Royal Society*, 464(2093):1141–1160, February 2008.

[14] C. Richard, J. C. M. Bermudez, and P. Honeine. Online prediction of time series data with kernels. *IEEE Transaction on Signal Processing*, 57(3):1058–1067, March 2009.

[15] M. T. M. Silva and V. H. Nascimento. Improving the tracking capability of adaptive filters via convex combination. *IEEE Transaction on Signal Processing*, 56(7):3137–3149, July 2008.

[16] W. Yan. Multiple model fusion for robust time-series forecasting. In *Time-Series Forecasting Workshop of International Joint Conference on Neural Networks (IJCNN)*, Orlando, FL, USA, August 12-17 2007.