We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



2

Reinforcement Learning in System Identification

Mariela Cerrada and Jose Aguilar
Universidad de los Andes
Mérida-VENEZUELA

1. Introduction

The Reinforcement Learning (RL) problem has been widely researched an applied in several areas (Sutton & Barto, 1998; Sutton, 1988; Singh & Sutton, 1996; Schapire & Warmuth, 1996; Tesauro, 1995; Si & Wang, 2001; Van Buijtenen et al., 1998). In dynamical environments, a learning agent gets rewards or penalties, according to its performance for learning good actions.

In identification problems, information from the environment is needed in order to propose an approximate system model, thus, RL can be used for taking the on-line information taking. Off-line learning algorithms have reported suitable results in system identification (Ljung, 1997); however these results are bounded on the available data, their quality and quantity. In this way, the development of on-line learning algorithms for system identification is an important contribution.

In this work, it is presented an on-line learning algorithm based on RL using the Temporal Difference (TD) method, for identification purposes. Here, the basic propositions of RL with TD are used and, as a consequence, the linear TD(λ) algorithm proposed in (Sutton & Barto, 1998) is modified and adapted for systems identification and the reinforcement signal is generically defined according to the temporal difference and the identification error. Thus, the main contribution of this paper is the proposition of a generic on-line identification algorithm based on RL.

The proposed algorithm is applied in the parameters adjustment of a Dynamical Adaptive Fuzzy Model (DAFM) (Cerrada et al., 2002; Cerrada et al., 2005). In this case, the prediction function is a non-linear function of the fuzzy model parameters and a non-linear TD(λ) algorithm is obtained for the on-line adjustment of the DAFM parameters.

In the next section the basic aspects about the RL problem and the DAFM are revised. Third section is devoted to the proposed on-line learning algorithm for identification purposes. The algorithm performance for time-varying non-linear systems identification is showed with an illustrative example in section fourth. Finally, conclusions are presented.

2. Theoretical background

2.1 Reinforcement learning and temporal differences

RL deals with the problem of learning based on trial and error in order to achieve the overall objective (Sutton & Barto, 1998). RL are related to problems where the learning agent does not know what it must do. Thus, the agent must discover an action policy for maximize the

Open Access Database www.i-techonline.com

Source: Reinforcement Learning: Theory and Applications, Book edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer
ISBN 978-3-902613-14-1, pp.424, January 2008, I-Tech Education and Publishing, Vienna, Austria

expected gain defined by the rewards that the agents gets. At time *t*, *(t=0, 1, 2, ...)*, the agent receives the *state* S_t and based on this information it choises an *action* a_t. As a consequence, the agent receives a *reinforcement signal or reward* r_{t+1}. In case of the infinite time domain, a *discount* weights the received reward and the *discounted expected gain* is defined as:

$$R_t = \sum_{k=0}^{\infty} \mu^k r_{t+k+1} \tag{1}$$

where μ, $0 < \mu \leq 1$, is the *discount rate*, and it determines the current value of the futures rewards.
On the other hand, TD method permits to solve the prediction problem taking into account the difference (error) between two prediction values at successive instants *t* and *t+1*, given by a function *P*. According to the TD method, the adjustment law for the parameter vector θ of the prediction function $P(\theta)$ in given by the following equation (Sutton, 1988) :

$$\theta_{t+1} = \theta_t + \eta \left[P(x_{t+1}, \theta_t) - P(x_t, \theta_t) \right] \frac{\partial P(x_t, \theta_t)}{\partial \theta} \tag{2}$$

where x_t is a vector of available data at time *t* and η, $0 < \eta \leq 1$, is the learning rate. The term between parentheses is the *temporal difference* and the equation (2) is the *TD algorithm* that can be used on-line in a incremental way.
RL problem can be viewed as a prediction problem where the objective is the estimation of the discounted gain defined by equation (1), by using the *TD* algorithm.
Let $\hat{R}_t$ be the prediction of R_t. Then, from equation (1):

$$R_t = r_{t+1} + \mu R_{t+1} \tag{3}$$

The real value of R_{t+1} is not available, then, by replacing it by its estimated value in (3), the prediction error is defined by the following equation:

$$\Delta = R_t - \hat{R}_t = r_{t+1} + \mu \hat{R}_{t+1} - \hat{R}_t \tag{4}$$

which describe a temporal difference. The reinforcement value r_{t+1} is defined in order to obtain at time *t+1* a better prediction of R_t, given by $\hat{R}_t$, based on available information. In this manner, a good estimation in the RL problem means the optimization of R_t.
Thus, denoting $\hat{R}$ as *P* and by replacing the temporal difference in (2) by that one defined in (4), the parameters adjustment law is:

$$\theta_{t+1} = \theta_t + \eta \left[r_{t+1} + \mu P(x_{t+1}, \theta_t) - P(x_t, \theta_t) \right] \frac{\partial P(x_t, \theta_t)}{\partial \theta} \tag{5}$$

The learning agent using the equation (5) for the parameters adjustment is called *Adaptive-Heuristic-Critic* (Sutton & Barto, 1998). In on-line applications, the time *t* is the same iteration time in the learning process by using equation (5).

2.2 Dynamical adaptive fuzzy models

Without loss of generality, a fuzzy logic model MISO (Multiple Inputs-Single Output), is a linguistic model defined by the following *M* fuzzy rules:

$$\theta_{t+1} = \theta_t + \eta \left[r_{t+1} + \mu P(x_{t+1}, \theta_t) - P(x_t, \theta_t) \right] \frac{\partial P(x_t, \theta_t)}{\partial \theta} \tag{6}$$

where x_i is a vector of linguistic input on the domain of discourse U_i ; y is the linguistic output variable on the domain of discourse V; F_i^l and G^l are fuzzy sets on U_i and V, respectively, *(i=1,... ,n)* and *(l=1,... ,M)*, each one defined by their membership functions.
The DAFM is obtained from the previous rule base (6), by supposing input values defined by fuzzy singleton, gaussian membership functions of the fuzzy sets defined for the fuzzy output variables and the defuzzification method given by center-average method. Then, the inference mechanism provides the following model (Cerrada et al., 2005):

$$y(X) = \frac{\sum_{l=1}^{M} \gamma^l \left(\prod_{i=1}^{n} \exp\left[-\frac{(x_i - \alpha_i^l)^2}{\beta_i^l} \right] \right)}{\sum_{l=1}^{M} \left(\prod_{i=1}^{n} \exp\left[-\frac{(x_i - \alpha_i^l)^2}{\beta_i^l} \right] \right)} \tag{7}$$

where $\underline{X}=(x_1\ x_2\ \ldots\ x_n)^T$ is a vector of linguistic input variables x_i at time t; $\alpha(v_i^l,t_j)$, $\beta(w_i^l,t_j)$ and $\gamma(u^l,t_j)$ are time-depending functions; v_i^l and w_i^l are parameters associated to the variable x_i in the rule l; u^l is a parameter associated to the center of the output fuzzy set in the rule l.
Definition. Let $x_i(t_j)$ be the value of the input variable x_i to the DAFM at time t_j to obtain the output $y(t_j)$. The generic structure of the functions $\alpha_i^l(v_i^l,t_j)$, $\beta_i^l(w_i^l,t_j)$ and $\gamma^l(u^l,t_j)$ in equation (7), are defined by the following equations (Cerrada et al., 2005):

$$\alpha_i^l = f(v_i^l, \bar{x}_i(t_j)) = v_i^l \frac{\sum_{k=j-\delta_1}^{j} x_i(t_k)}{\delta_1 + 1} \quad \delta_1 \in N \tag{8}$$

$$\beta_i^l = f(w_i^l, \sigma_i^2(t_j)) = w_i^l \left(\frac{\sum_{k=j-\delta_1}^{j} (x_i(t_k) - \bar{x}_i(t_k))^2}{\delta_1 + 1} + \varepsilon \right), \quad \varepsilon \in \Re, \delta_1 \in N, \bar{x}_i(t_o) = x_i(0) \tag{9}$$

$$\gamma^l = f(u^l, \overline{y}(t_j)) = u^l \overline{y}(t_j) \tag{10}$$

where:

$$\overline{y}(t_j) = \frac{\sum_{k=j-\delta_2}^{j-1} y(t_k)}{\delta_2}, \quad \delta_2 \in N; \quad y(t_O) = y(0) \tag{11}$$

or

$$\overline{y}(t_j) = \frac{\sum_{k=1}^{j-1} y(t_k)}{j-1}, \; y(t_o) = y(0) \tag{12}$$

The parameters v_i^l, w_i^l and u^l can be on-line or off-line adjusted by using the following iterative generic algorithm:

$$\theta(k+1) = \theta(k) + \eta\Delta_\theta \tag{13}$$

where $\theta(t)$ denotes the vector of parameters at time t, Δ_θ is the parameter increment at time t and η, $0<\eta<1$, is the learning rate. Tuning algorithm by using off-line gradient-based learning is presented in (Cerrada et al., 2002; Cerrada et al., 2005).
In this work, the initial values of parameters are randomly selected on certain interval, the number of rules M is fixed and it is not adjusted during the learning process. The input variables x_i are also known, then, the number of adjustable parameters is fixed.
Clearly, by taking the functions $\alpha_i^l(v_i^l,t_j)$, $\beta_i^l(w_i^l,t_j)$ and $\gamma^l(u^l,t_j)$ as parameters in equation (7), a classical Adaptive Fuzzy Model (AFM) is obtained (Wang, 1994). The mentioned parameters are also adjusted by using the learning algorithm (13). Comparisons between the performances of the AFM and DAMF in system identification are provided in (Cerrada et al., 2005).

3. RL-based on-line identification algorithm

In this work, the fuzzy identification problem is solved by using the weighted identification error as a prediction function in the RL problem, and by suitably defining the reinforcement value according to the identification error. Thus, the minimization of the prediction error (4) drives to the minimization of the identification error.
The *critic* (learning agent) is used in order to predict the performance on the identification as an approximator of the system's behavior. The prediction function is defined as a function of the *identification error* $e(t,\theta_t)=y(t)-y_e(t,\theta_t)$, where $y(t)$ denotes the real value of the system output at time t and $y_e(t,\theta_t)$ denotes the estimated value given by the identification model by using the available values of θ at time t.
Let P_t be the proposed non-linear prediction function, defined as a cumulative addition on an interval of time, given by the following equation :

$$P(x_t,\theta_t) = \frac{1}{2}\sum_{k=t-K}^{t} (\mu\lambda)^{t-k}\, e^2(x_k,\theta_t) \tag{14}$$

where $e(x_k,\theta_t)=y(k)-y_e(x_k,\theta_t)$ defines the identification error at time k and the value of θ at time t, and K defines the size of the time interval. Then:

$$\frac{\partial P(x_t,\theta_t)}{\partial\theta} = \sum_{k=t-K}^{t} (\mu\lambda)^{t-k}\, e(x_k,\theta_t)\Psi(k,\theta_t) \tag{15}$$

where:

$$\Psi(k,\theta_t)=\frac{\partial e(x_k,\theta_t)}{\partial\theta}=-\frac{\partial y_e(k,\theta_t)}{\partial\theta} \tag{16}$$

By replacing (15) into (5), the following learning algorithm for the parameters adjustment is obtained:

$$\theta_{t+1}=\theta_t+\eta\left[r_{t+1}+\mu P(x_{t+1},\theta_t)-P(x_t,\theta_t)\right]\sum_{k=t-K}^{t}(\mu\lambda)^{t-k}e(x_k,\theta_t)\Psi(k,\theta_t) \tag{17}$$

where expression in equation (15) can be viewed as the *eligibility trace* (Sutton & Barto, 1998), which stores the temporal record of the identification errors weighted by the parameter λ. From (14), the function $P(x_{t+1},\theta_t)$ is obtained in the following manner:

$$\begin{aligned}P(x_{t+1},\theta_t)&=\frac{1}{2}\sum_{k=t-K}^{t+1}(\mu\lambda)^{t+1-k}e^2(x_k,\theta_t)\\&=\frac{1}{2}\left[e^2(x_{t+1},\theta_t)+\sum_{k=t-K}^{t}(\mu\lambda)^{t+1-k}e^2(x_k,\theta_t)\right]\\&=\frac{1}{2}e^2(x_{t+1},\theta_t)+\mu\lambda\left[\frac{1}{2}\sum_{k=t-K}^{t}(\mu\lambda)^{t-k}e^2(x_k,\theta_t)\right]\\&=\frac{1}{2}e^2(x_{t+1},\theta_t)+\mu\lambda P(x_t,\theta_t)\end{aligned} \tag{18}$$

By replacing (18) into (17), the learning algorithm is given.
In the prediction problem, a good estimation of R_t is expected; that implies $P(x_t,\theta_t)$ goes to $r_{t+1}+\mu P(x_{t+1},\theta_t)$. This condition is obtained from equation (4). Given that the prediction function is the weighted sum of the square identification error $e^2(t)$, then it is expected that:

$$0\le r_{t+1}+\mu P(x_{t+1},\theta_t)<P(x_t,\theta_t) \tag{19}$$

On the other hand, a suitable adjustment of identification model means that the following condition is accomplished:

$$0\le P(x_{t+1},\theta_t)<P(x_t,\theta_t) \tag{20}$$

The reinforcement r_{t+1} is defined in order to accomplish the expected condition (19) and taking into account the condition (20). Then, by using equations (14) and (18), the reinforcement signal is defined as:

$$r_{t+1}=-\frac{1}{2}\mu e^2(x_{t+1},\theta_t) \quad \text{if} \quad P(x_{t+1},\theta_t)>P(x_t,\theta_t) \tag{21}$$

$$r_{t+1}=0 \quad \text{if} \quad P(x_{t+1},\theta_t)\le P(x_t,\theta_t) \tag{22}$$

In this way, the identification error into the prediction function $P(x_{t+1},\theta_t)$, according to the equation (18), is rejected by using the reinforcement in equation (22). The learning rate η in (17) is defined by the following equation:

$$\eta(k) = \frac{\eta(k-1)}{\rho + \eta(k-1)}; \quad 0 < \rho < 1 \tag{23}$$

Thus, an accurate adjustment of parameters is expected. Usually, $\eta(0)$ is around *1*, and ρ is around *0*. Parameters μ and λ can depend on the system dynamic: small values in case of slow dynamical systems, and values around *1* in case of fast dynamical systems.
In this work, the proposed RL-based algorithm is applied to fuzzy identification and the identification model is provided by the DAFM in (7). Then, the prediction function P is a non-linear function of the fuzzy model parameters and a non-linear approach of $TD(\lambda)$ is obtained.

3.1 Descent-gradient-based analysis

The proposed identification learning algorithm can be studied like a descent-gradient method with respect to the parametric predictive function P. In the descent-gradient method for optimization, the objective is to find the minimal value of the error measure on the parameters space, denoted by $J(\theta)$, by using the following algorithm for the parameters adjustment:

$$\theta_{t+1} = \theta_t + \Delta_{\theta_t} = \theta_t + 2\alpha\left[\mathrm{E}\{z \mid x_t\} - P(x_t, \theta)\right]\nabla_\theta P(x_t, \theta) \tag{24}$$

In this case, a error measure is defined as:

$$J(\theta, x) = \left(\mathrm{E}\{z \mid x\} - P(x, \theta)\right)^2 \tag{25}$$

where $E\{z \mid x\}$ is the expected value of the real value z, from the knowledge of the available data x.
In this work, the learning algorithm (17) is like a learning algorithm (24), based on the descent-gradient method, where $r_{t+1} + \mu P(x_{t+1}, \theta_t)$ is the expected value $E\{z \mid x\}$ in (25). By appropriate selecting r_{t+1} according to (21) and (22), the expected value in the learning problem is defined in two ways:

$$\mathrm{E}\{z \mid x\} = \mu P(x_{t+1}, \theta_t) \quad \text{if} \quad P(x_{t+1}, \theta_t) \le P(x_t, \theta_t) \tag{26}$$

or

$$\mathrm{E}\{z \mid x\} = \mu^2 \lambda P(x_t, \theta_t) \quad \text{if} \quad P(x_{t+1}, \theta_t) > P(x_t, \theta_t) \tag{27}$$

Then, the parameters adjustment is made on each iteration in order to attain the expected value of the prediction function P according to the predicted value of $P(x_{t+1}, \theta_t)$ and the real value $P(x_t, \theta_t)$. In both of cases, the expected value is minor than the obtained real value $P(x_t, \theta_t)$ and the selected value of r_{t+1} defines the magnitude of the defined error measure.

4. Illustrative example

This section shows an illustrative example applied to fuzzy identification of time-varying non-linear systems by using the proposed on-line RL-based identification algorithm and the DAFM described in section 2.2. Comparisons by using off-line gradient-based tuning

algorithm are presented in order to highlight the algorithm performance. For off-line adjustment purposes, the input-output training data is obtained from Pseudo-Random Binary Signal (PRBS) input signal. The performance of the fuzzy identification is evaluated according to the identification relative error $(e_r=(y(t)-y_e(t))/y(t))$ normalized on $[0,1]$.
The system is described by the following difference equation:

$$y(k+1) = g[y(k), y(k-1)] + u(k) \tag{28}$$

where

$$g[y(k), y(k-1)] = \frac{y(k)y(k-1)[y(k)+a(k)]}{1+y^2(k)+y^2(k-1)} \tag{29}$$
$$a(k) = 2.5 + 2.5\ sen(2\pi k/250)$$

In this case, the unknown function $g=[.]$ is estimated by using the DAFM and, additionally, a sudden change on $a(k)$ is proposed by setting $a(k)=0$, $k>400$. After an extensive training phase, the fuzzy model with $M=8$, $\delta_1=4$ and $\delta_2=1$ (in equations (8),(9),(11)), has been chosen. In this case, the fuzzy identification performance is adequate and the Root Mean Square Error (RMSE) is 0.1285 in validation phase. Figure 1 shows the performance of the DAFM using the off-line gradient-based tuning algorithm with initial conditions on the interval $[0,1]$ and using the following input signal:

$$u(k) = \begin{cases} sen(2\pi k/25) & 1 < k < 100 \quad \text{y} \quad k > 500 \\ 3 + (0.5sen(2\pi k/250) + 0.5sen(2\pi k/25)) & 100 < k < 500 \end{cases} \tag{30}$$

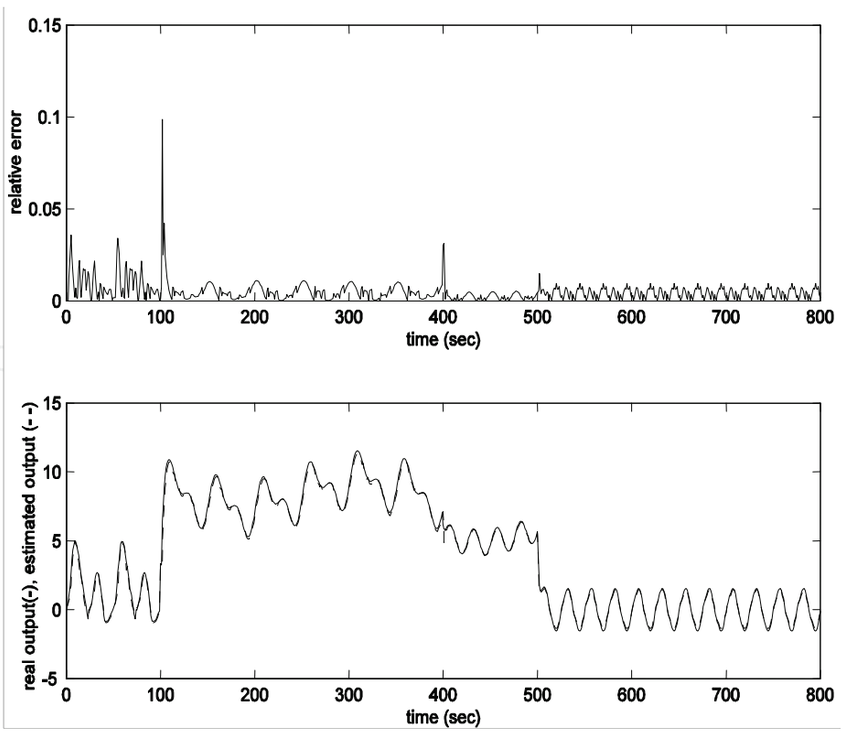


Fig. 1. Fuzzy identification using off-line tuning algorithm and DAFM

In the following, fuzzy identification performance by using the DAFM with the proposed RL-based tuning algorithm is presented. Equation (17) is used for the parameters adjustment with the prediction function defined in (14) and the reinforcement defined in (21)-(22). Here, $\lambda=\mu=0.9$, $K=5$ and the learning rate is set up by the equation (23) with $\rho=0.01$. Note that the iteration index t is the same time k in system (28). After experimental proofs, the performance approaching the accuracy obtained from off-line adjustment is obtained with $M=6$ and initial conditions on $[0.5,1.5]$. Here, the RMSE= 0.0838 is achieved. Figure 2 shows the tuning algorithm performance and table 1 shows the comparative values related to the RMSE.

M	RMSE off-line	RMSE On-line
6	0.1110	0.0838
8	0.1285	0.1084
10	0.1327	0.1044
15	0.1069	0.0860
20	0.1398	0.1056

Table 1. Comparison between the on-line proposed algorithm and off-line tuning

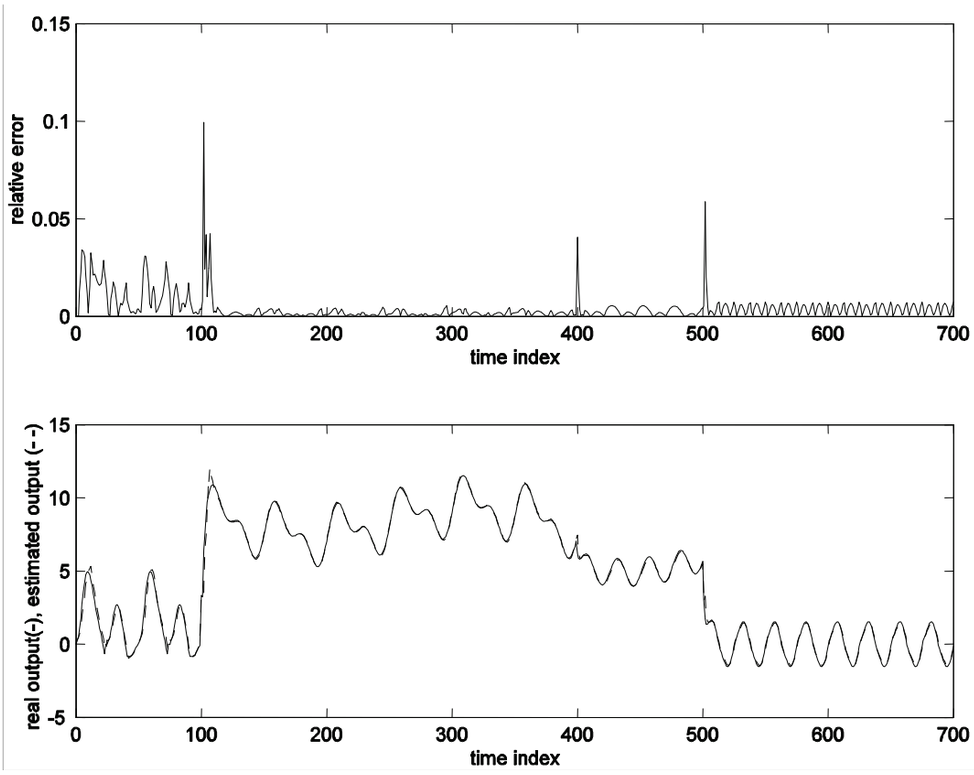


Fig. 2. Fuzzy identification using RL-based tuning algorithm and DAFM

4.1 Initial condition dependence

In order to show the algorithm sensibility according to the initial conditions of the fuzzy model parameters, the following figures show the tuning algorithm performance. In this case, the system is described by the equation (31):

$$y(k+1) = \frac{y(k)y(k-1)y(k-2)u(k-1)(y(k-2)-1)+u(k)}{a(k)+y(k-2)^2+y(k-1)^2} \tag{31}$$

$$a(k) = 1 + 0.1\ sen(2\pi k/100)$$

where:

$$u(k) = \begin{cases} sen(2\pi k/250) & 1 < k < 500 \text{ and } 801 < k < 1000 \\ 1.5 + (0.8sen(2\pi k/250) + 0.2sen(2\pi k/25)) & 501 < k < 800 \end{cases} \tag{32}$$

Figure 3 shows the tuning process by using a model with *M=20* and initial conditions on the interval *[0.5,1.5].* In this case, even when the initial error is large, the tuning algorithm performance also shows an adequate performance and the tuning process has an suitable evolution (here, a sudden change on *a(k)* is not considered). Figure 4 shows the tuning process by using a model with initial conditions on the interval *[0,1]* an also a suitable performance of the proposed identification algorithm is shown.

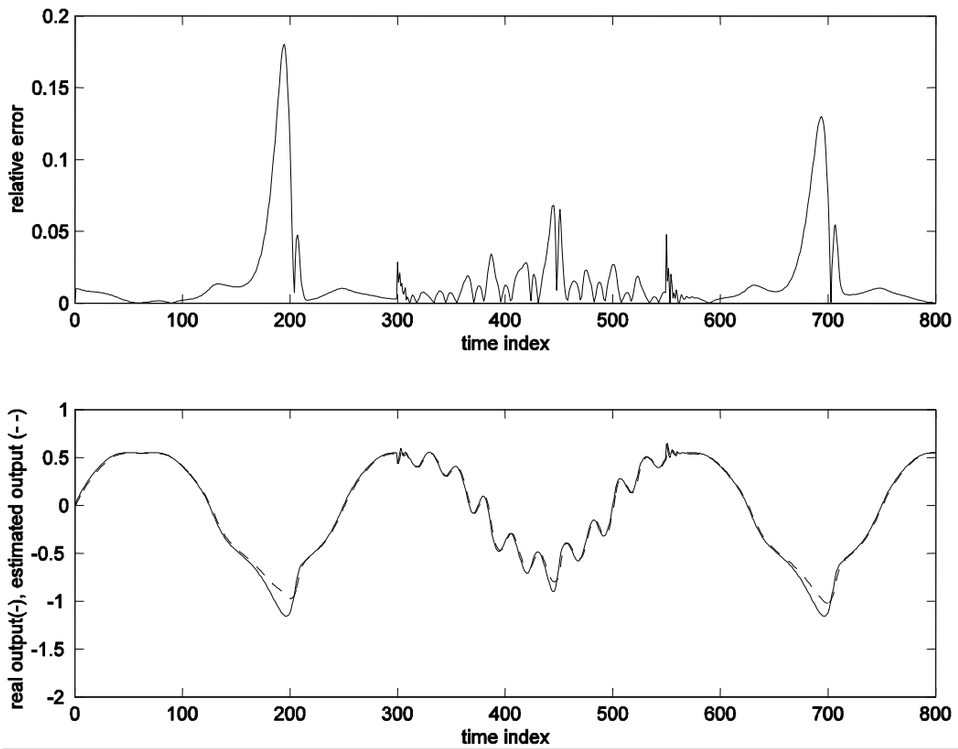


Fig. 3. Fuzzy identification using RL-based tuning algorithm and DAFM with initial conditions on *[0.5,1.5].*

The previous tests show the performance and the sensibility of the proposed on-line algorithm is adequate in terms of (a) The initial conditions of the DAFM parameters, (b) Changes on the internal dynamic (the term *a(k)* in the example) and (c) Changes on the inputs signal (the proposed input *u(k)*).
These ones are very important aspects to be evaluated in order to consider an on-line identification algorithm. In the example, even though the initial error depends on the initial conditions of the DAFM parameters, a good evolution of the learning algorithm is accomplished. Table 1 also shows the number of rules *M* do not strongly determines the global performance of the proposed on-line algorithm although a similar RMSE could be obtained with a low number of rules and off-line tuning. However, this one could be not reached whether good quality and quantity of historical data is not available in off-line approaches.

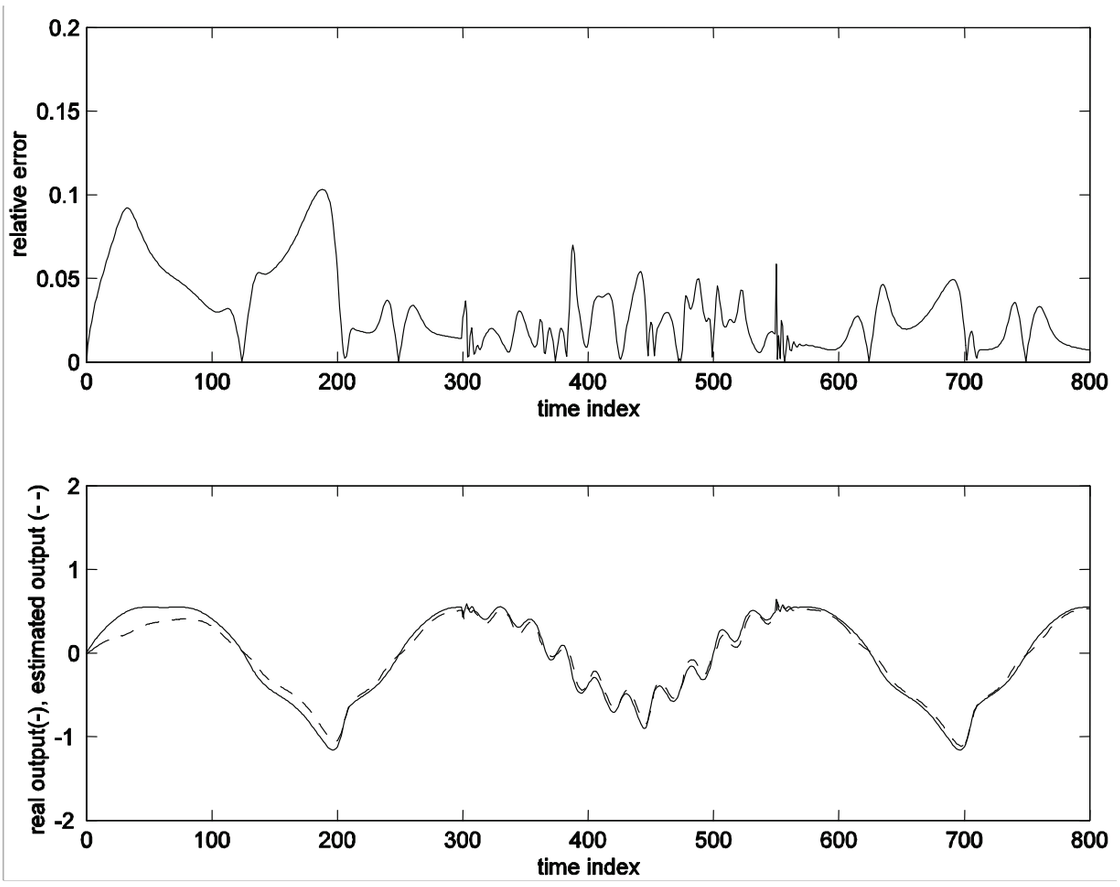


Fig. 4. Fuzzy identification using RL-based tuning algorithm and DAFM with initial conditions on *[0,1].*

5. Acknowledgment

This work has been partially presented in the International Symposium on Neural Networks ISNN 2006.

6. Conclusion

This work proposes an on-line tuning algorithm based on reinforcement learning for the identification problem. Both the prediction function and the reinforcement signal have been defined by taking into account the identification error, according to the classical recursive identification algorithms. The presence of the reinforcement signal in the proposed tuning algorithm permits to reject the identification error into the prediction function, then, the parameters adjustment not only depends on the gradient direction.

The proposed algorithm has been applied in fuzzy identification, then, the prediction function is a non-linear function of the fuzzy model parameters. In this case, the proposed identification model is a Dynamical Adaptive Fuzzy Model (DAFM) that has reported a good performance in identification problems.

In order to show the algorithm performance, an illustrative example related to time-varying non-linear system identification using a DAFM has been developed. The obtained results have been compared by using the off-line gradient-based learning algorithm. The performance obtained by using the DAFM with the proposed on-line algorithm is adequate in terms of the main aspects to be taken into account in on-line identification: the initial conditions of the model parameters, the changes on the internal dynamic and the changes on the input signal.

Even when similar results could be obtained by using the DAFM with off-line tuning, in this case good quality and quantity of available historical data is needed to reach a suitable validation phase in off-line tuning. This one highlights the use of the on-line learning algorithms and the proposed RL-based on-line tuning algorithm could be an important contribution for the system identification in dynamical environments with perturbations, for example, in process control area.

7. References

Cerrada, M.; Aguilar, J.; Colina, E. & Titli, A. (2002). An approach for dynamical adaptive fuzzy modeling, *Proceedings of IEEE-FUZZ 2002 International Conference on Fuzzy Systems*, pp. 156-161, Hawai-USA, May 2002, Canada

Cerrada, M.; Aguilar, J.; Colina, E. & Titli, A. (2005). Dynamical membership functions: an approach for adaptive fuzzy modelling. *Fuzzy Sets and Systems*, Vol. 152, No. 3, (June 2005) (513-533)

Ljung, L. (1997). *System Identification. Theory for the User,* Prentice Hall, New York

Schapire, R.E. & Warmuth, M.K. (1996). On the worst-case analysis of temporal differences learning algorithms. *Machine Learning,* Vol. 22, (95-121)

Si, J. & Wang, Y-T. (2001). On line learning control by association and reinforcement. *IEEE Transactions on Neural Networks*, Vol. 12, No. 2, (264-276)

Singh, S.P. & Sutton, R.S. (1996). Reinforcement learning with replacing eligibility traces. *Machine Learning*, Vol. 22, (123-158)

Sutton, R.S. & Barto, A.G. (1998). *Reinforcement Learning. An Introduction,* The MIT Press, Cambridge

Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning,* Vol. 3, (9-44)

Tesauro, G. (1995). Temporal difference learning and TD-Gammon. *Communications of the Association for Computing Machinery*, Vol. 38, No. 3, (58-68)

Van Buijtenen, W.M.; Schram, G.; Babuska, R. & Verbruggen, H.B. (1995). Adaptive fuzzy control of satellite attitude by reinforcement learning. *IEEE Transactions on Fuzzy Systems*, Vol. 6, No. 2, (185-194)

Wang, L.X. (1994). *Adaptive Fuzzy Systems and Control,* Prentice Hall, New Jersey

IntechOpen



Reinforcement Learning

Edited by Cornelius Weber, Mark Elshaw and Norbert Michael Mayer

ISBN 978-3-902613-14-1
Hard cover, 424 pages
Publisher I-Tech Education and Publishing
Published online 01, January, 2008
Published in print edition January, 2008

Brains rule the world, and brain-like computation is increasingly used in computers and electronic devices. Brain-like computation is about processing and interpreting data or directly putting forward and performing actions. Learning is a very important aspect. This book is on reinforcement learning which involves performing actions to achieve a goal. The first 11 chapters of this book describe and extend the scope of reinforcement learning. The remaining 11 chapters show that there is already wide usage in numerous fields. Reinforcement learning can tackle control tasks that are too complex for traditional, hand-designed, non-learning controllers. As learning computers can deal with technical complexities, the tasks of human operators remain to specify goals on increasingly higher levels. This book shows that reinforcement learning is a very dynamic area in terms of theory and applications and it shall stimulate and encourage new research in this field.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mariela Cerrada and Jose Aguilar (2008). Reinforcement Learning in System Identification, Reinforcement Learning, Cornelius Weber, Mark Elshaw and Norbert Michael Mayer (Ed.), ISBN: 978-3-902613-14-1, InTech, Available from:
http://www.intechopen.com/books/reinforcement_learning/reinforcement_learning_in_system_identification

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2008 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen