VERSITA

Print ISSN:1312 – 2622; Online ISSN: 2367-5357
DOI:10.2478/itc-2013-0004

Reinforcement Learning for Predictive Maintenance of Industrial Plants

P. Koprinkova-Hristova

***Key Words**: Reinforcement learning; temporal difference error; echo state network; predictive maintenance; industrial plant.*

***Abstract**. The reinforcement learning is a well-known approach for solving optimization problems having limited information about plant dynamics. Its key element, named "critic" is aimed at prediction of future "punish/reward" signals received as a result of undertaken control actions. The main idea in the present work is to use such a "critic" element for prediction of approaching alarm situations based on limited measurement information from the industrial plant. In order to train the critic network in real time it is proposed to use a special kind of a fast trainable recurrent neural network, called Echo State Network (ESN). The approach proposed is demonstrated on an example for predictive maintenance of a mill fan in Maritsa East 2 Thermal Power Plant.*

Introduction

Reinforcement learning (RL) aroused as a method for training of artificial neural networks "by experience", rather than "by examples". It was created to mimic animals' behaviour in an attempt to explain Pavlovian conditioning [1]. Later on RL was also recognized as an approximation of Bellman's dynamic programming method [2], that is well-known in control community. During the last thirty years the theoretical developments in this field (a very detailed retrospective can be found in [12]) have led to methodologies known as Neuro Dynamic Programming (NDP) [3] and Adaptive Critic Designs (ACD) [14], also commonly known as Adaptive Dynamic Programming (ADP).

The core of this group of methods is the approximation of Bellman's equation for the current time instant, using a neural network called "heuristic adaptive critic". The result obtained by such a "critic" prediction is named a "value function". Originally it was the discounted sum of future values of a binary "punish/reward" signal coming from the animals' environment as a result of some actions undertaken by them. Later it was replaced by a more complicated utility function representing the outcomes or losses that have to be maximized or minimized respectively. The predicted result by the well trained critic "value" is used to optimize the controller or "actor" behaviour in terms of RL. Different training schemes for adaptive critic design in dependence on the presence or absence of a model of the plant [14, 16] were developed. In both cases the critic was trained, using Temporal Difference (TD) error [17], thereby mimicking the brain's ability to learn how to predict future outcomes on the basis of previous experience without awaiting the results of the future actions.

The main aim of the predictive maintenance of industrial plants is prediction on time of any undesirable future regimes, based on the available story of the on-line measurements of some key variables. If an exact model of the plant is available, it could be easily used for prediction of its future states and hence, the current working regime of the plant. However, real industrial plant identification by an adequate model is usually a hard task. In spite of this, experienced plant operators are able to predict future alarm situations based only on available real time measurements. Similarly, in terms of RL, a well trained critic is able to predict future rewards or losses without any model of the environment and using only some sensory inputs. This analogy provoked the approach presented in this paper: to train an adaptive critic able to predict on time the approaching alarm conditions of an industrial plant using only the available measurements information, without the need of an adequate plant model. The approach suggested is demonstrated on an example for predictive maintenance of a mill fan in Maritsa East 2 Thermal Power Plant.

Adaptive Dynamic Programming and Reinforcement Learning

The common optimization task is defined as follows: for a given discrete dynamical plant described by the model $S(k+1)=F(S(k), a(k)), k=0,1,\ldots N–1$, find such control policy $a(1), a(2), \ldots, a(N–1)\}$, that maximizes (minimizes) a given utility function $U(k)=G(S(k), a(k)), k=0,1,\ldots N–1$ at all time instants k. Here $S(k)$ denotes the current plant state and $a(k)$ – the currently applied action.

The dynamic programming method proposes the following decision of this optimization task: maximize (minimize) the discounted sum of the utility function

$$(1)\qquad J(S(t), a(k)) = \sum_{k=0}^{N} \gamma^k U(S(t+k), a(t+k))$$

by solving N maximization (minimization) sub-tasks backwards starting from the final time instant to the initial one as follows:

$$(2)\qquad J(S(N), a(N)) = U(S(N), a(N)) \to \quad \max(\min)$$

$$(3)\qquad J(S(k), a(k)) = \max_{a(k)\in A}(\min)\{U(S(k), a(k)) + \gamma J(S(k+1))\}$$

Here $0<\gamma<1$ is called a discount factor. The last equation is the well-known Bellman's equation.

Created to overcome the well-known "curse

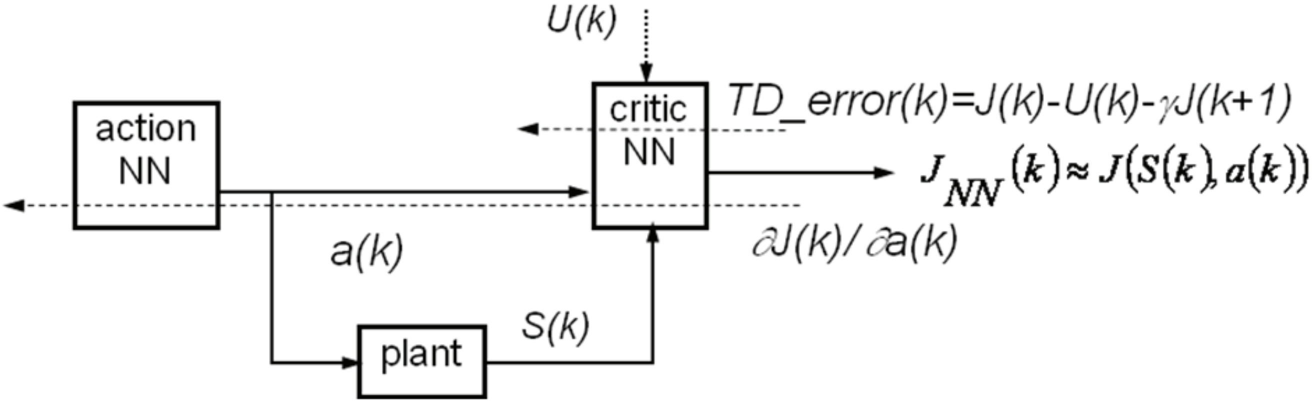
U(k)
action NN
critic NN
TD_error(k)=J(k)-U(k)-γJ(k+1)
$J_{NN}(k) \approx J(S(k), a(k))$
∂J(k)/∂a(k)
a(k)
plant
S(k)

Figure 1. Adaptive critic design

of dimensionality" of this approach, Adaptive Critic Design (ACD), which basic scheme is shown in *figure 1*, proposes approximation of Bellman's equation using some non-linear dependence (usually a neural network model) $J(S(k), a(k)) \approx J_{NN}(k)$. This approximation model is called an "adaptive critic" or briefly "critic".

Obtaining such model allows to solve the optimization task in a forward manner by maximizing (minimizing) the currently predicted discounted sum of future utilities *J*. The adjustment of the action network (the dashed lines in *figure 1*) is done by a gradient method using a critic network to calculate the derivatives of *J* with respect to the actions applied to the plant.

Here the focus is only on training of the critic neural network that is able to predict the current value function *J*. The training of such NN critic is aimed at minimization of the temporal difference error that comes from the postulation that the critic must approximate the right side of Bellman's equation (3), i.e.

(4) $$TD\_error(k) = J_{NN}(k) - U(k) - \gamma J_{NN}(k+1)$$

Hence, the square error that has to be minimized is

(5) $$\|E(k)\| = TD\_error(k)^2$$

Adaptive Critic Training for Predictive Maintenance

For the purpose of predictive maintenance it is proposed here to use only the critic part from the overall ACD scheme (*figure 2*). The dashed lines represent the information flow of the parameter tuning algorithms for the critic network. The dotted-line arrow for the utility function intends to show that this is not the actual input to the critic NN, but it is only the information supplied from the external environment (or experienced operator) needed to calculate the current temporal difference error. The proposal here is to replace the utility function $U(k)$ by the current working regime of the plant. This working regime is determined on the basis of the plant current state using some expert information about its meaning to the technologists. It can be defined by a linguistic variable having linguistic values like *satisfactory*, *good*, *deteriorating* or *critical*. These linguistic values can be associated with consecutive numbers starting from "0" for regimes that are far from the critical point and increasing for regimes approaching any emergency situations of the plant. For example, the above described four linguistic variables can be associated with the numbers called "cases" *C* as shown in *table 1*.

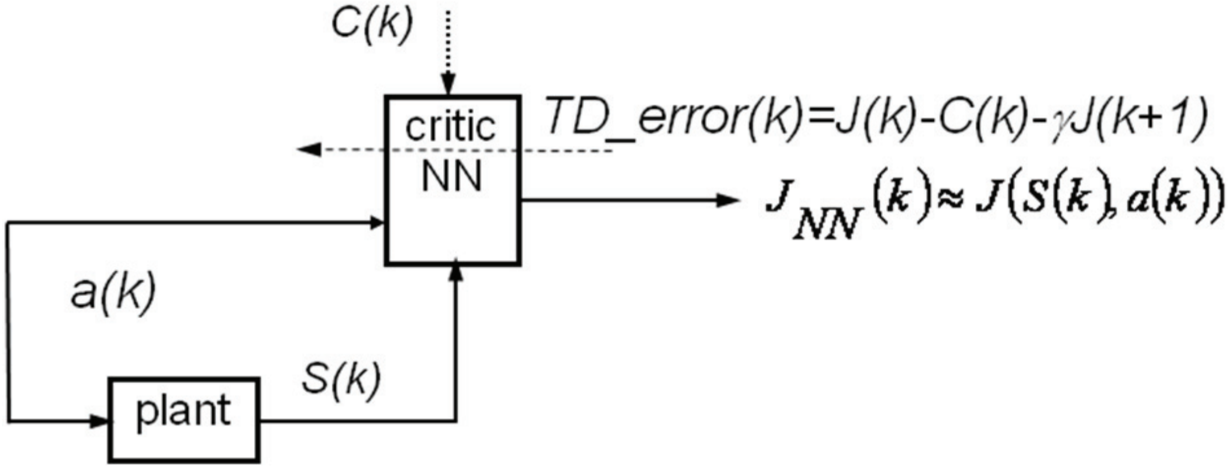
C(k)
critic NN
TD_error(k)=J(k)-C(k)-γJ(k+1)
$J_{NN}(k) \approx J(S(k), a(k))$
a(k)
plant
S(k)

Figure 2. Adaptive critic training scheme

Table 1. Plant working regimes defined as cases

Working regime	**Cace number *C***
Good	0
Satisfactory	1
Deteriorating	2
Critical	3

Hence, the bigger the case number is, the closer the probable plant alarm situation is. Then if the utility $U(k)$ is replaced by the current case $C(k)$, we can train a critic NN able to predict deterioration of the plant working regime for predictive maintenance purposes, i.e.

$$J(S(t),a(k)) = \sum_{k=0}^{N} \gamma^k C(S(t+k),a(k)) \quad (6)$$

The plant state $S(k)$ could be replaced by a vector of all available real-time measurements from the plant.

The advantages of such an approach are:

a) It does not need an adequate plant model to predict its future state.

b) Using of NN critic allows its on-line adaptation, e.g. by introduction of a feedback by an experienced plant operator about the current plant regime or when some new information is available.

Echo State Networks (ESN)

Since real industrial plants are non-linear and with time delays, approximation of the utility function (and its discounted sum *J*) will need a dynamic neural network, i.e., Recurrent Neural Networks (RNN) are proper candidates. However, their training in real time is prevented because the more complicated algorithms work slowly. That is why in the present study the use of a recently developed fast trainable RNN structure, called Echo State Network (ESN) is proposed. ESNs are a kind of recurrent neural networks that arise from the so called "reservoir computing approaches" [13]. The basic ESN structure is shown in *figure 3*.

The ESN output vector $J(k)$ for the current time instance k is usually a linear function of its input and current state

$$J(k) = f^{out}\left(W^{out}\left[in(k),\ \ X(k)\right]\right)$$

Here $in(k)$ is a vector of the network inputs and $X(k)$ is a vector composed of the reservoir neural states; f^{out} is a linear function (usually the identity), W^{out} is a trainable $n_y \times (n_{in} + n_X)$ matrix (here n_y, n_{in} and n_X are the sizes of the corresponding vectors y, u and X). The neurons in the reservoir have a simple sigmoid output function f^{res} (usually *tanh*) that depends on both the ESN input $u(k)$ and the previous reservoir state $X(k-1)$

$$X(k) = f^{res}\left(W^{in} in(k) + W^{res} X(k-1)\right).$$

W^{in} and W^{res} are $n_{in} \times n_X$ and $n_X \times n_X$ matrices that are randomly generated and are not trainable. There are different approaches for reservoir parameter production [13]. A recent approach used in the present investigation is proposed in [15]. It is called Intrinsic Plasticity (IP) and suggests initial adjustment of these matrices, aiming at increasing the entropy of the reservoir neurons outputs. ESN training can be done in an off-line or an on-line mode. For on-line training, the RLS algorithm [5] was proposed.

Mill Fan Working Regimes Prediction

The proposed approach is tested on an example for prediction of the current working regime ("case") of a mill fan in Maritsa East 2 Thermal Power Plant. The mill fans are a basic element of the dust-preparing systems of steam generators with direct breathing of the coal dust in the furnace chamber. They are a part of the equipment of power units that are most often repaired due to intensive erosion of the operative wheel blades in the process of grinding of low-calorific lignite coal.

In the previous works [4,9] the data archived by the installed on site Decentralized Control System of Maritsa East 2 about a mill fan rotor vibrations (*V*), dust-air mixture temperature (*T*) and the plant controller action (*A*) were analyzed. The observation period is 01.06.2010–06.11.2010.

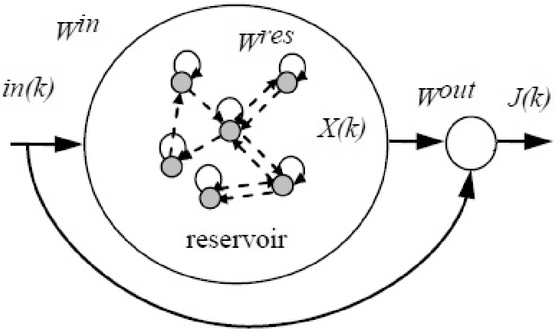

Figure 3. Echo state network structure

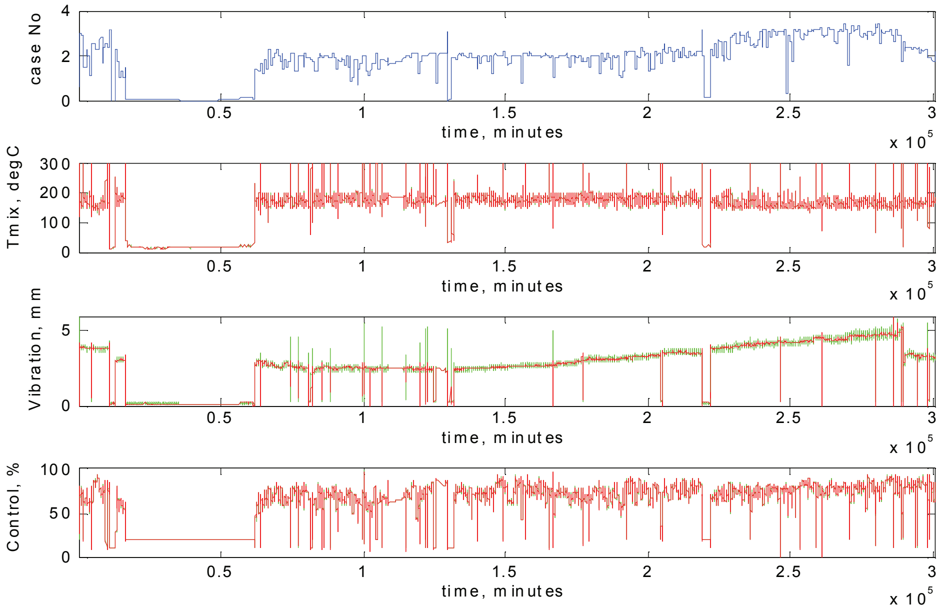

Figure 4. Fuzzy classification of the process state by the trained Sugeno fuzzy rule base using "blind" clistering approach

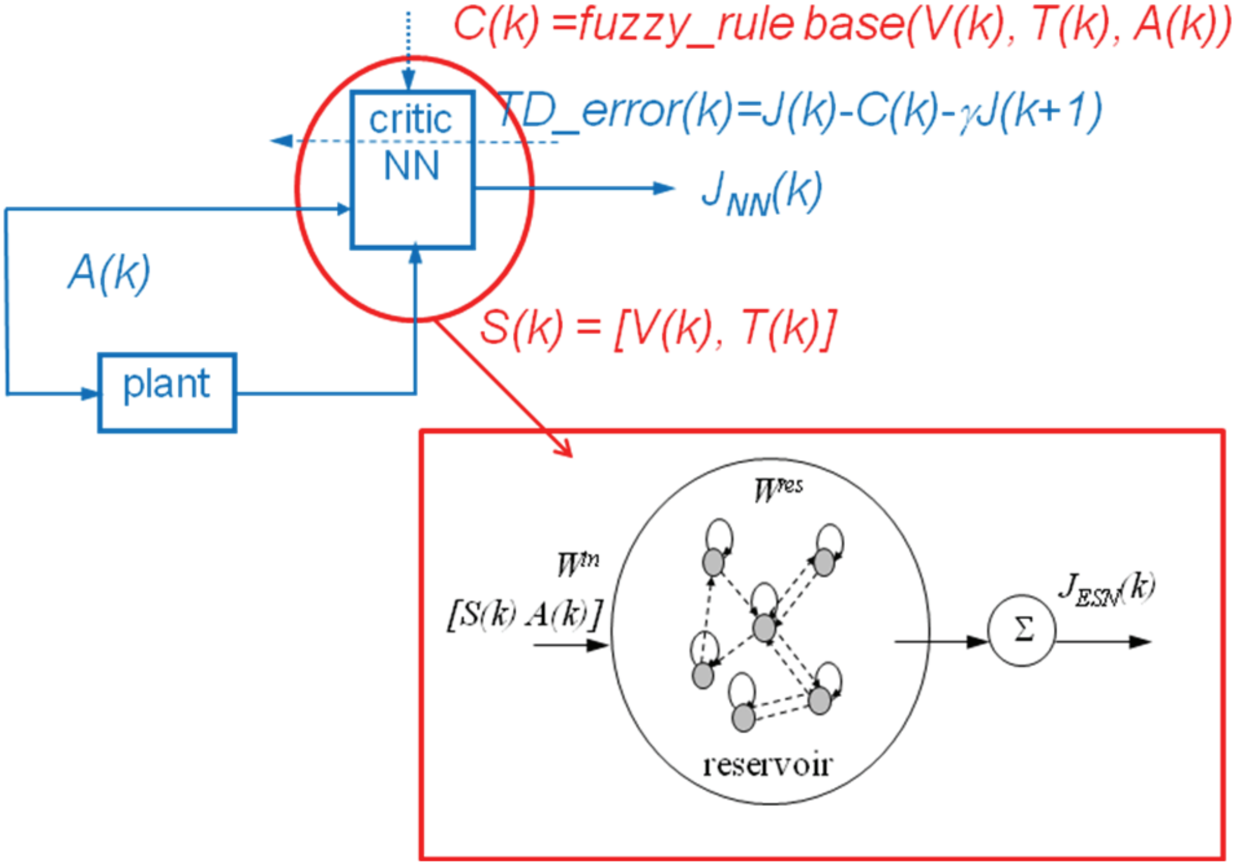

Figure 5. ESN critic for predictive maintenance

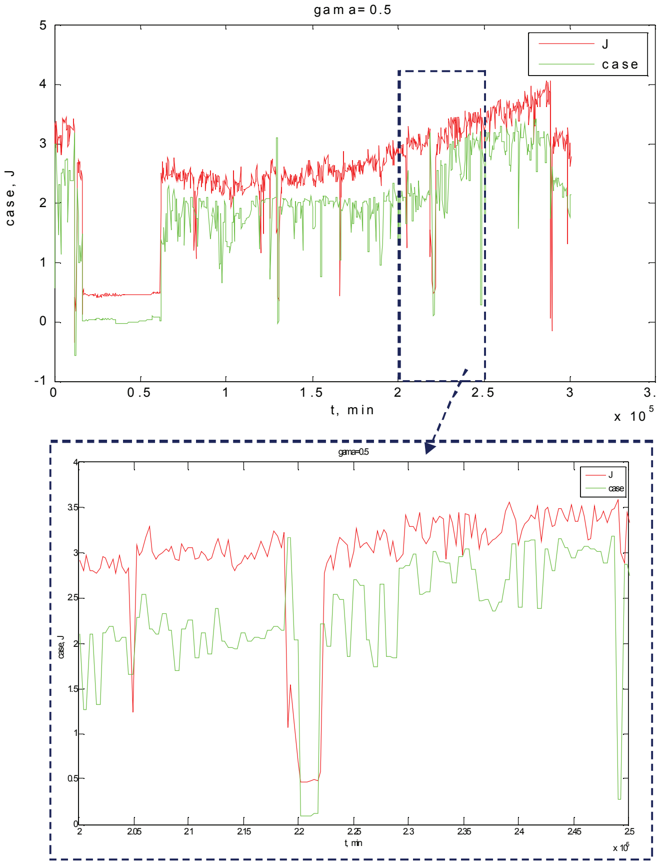

Figure 6. ESN case predictor for γ=0.5 and "zoom" for a chosen period of time

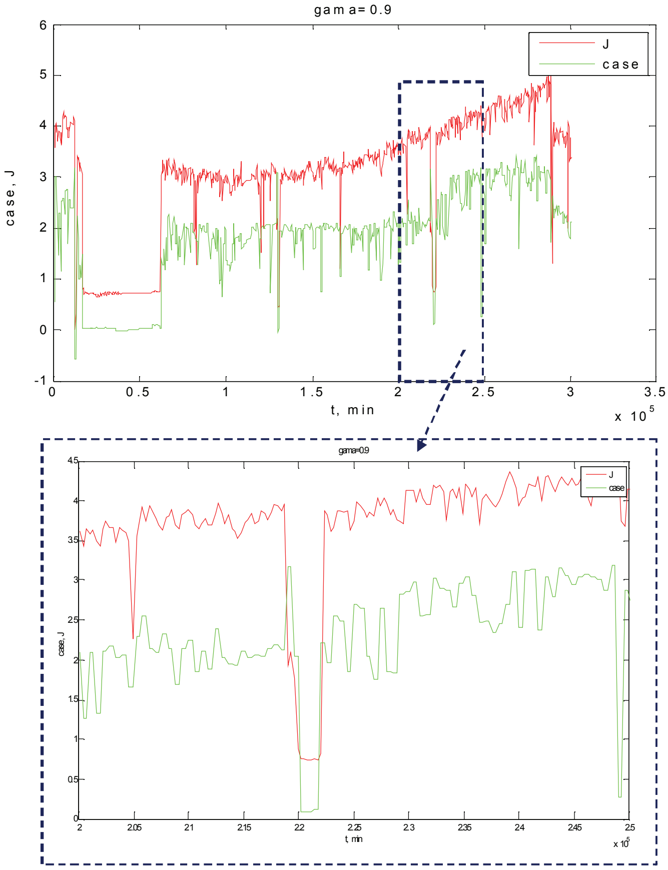

Figure 7. ESN case predictor for γ=0.9 and "zoom" for a chosen period of time

Table 2. Mill fan working regimes defined as cases

Working regime	**Cace number *C***
Stopping	0
Starting	1
Stable work	2
Deterioration	3

The period chosen allows process analyses before (the period 01.06–31.10.2010) and after (the period 01–06.11.2010) the replacement. All measurements are taken at 1 min intervals. Based on the available expert information and on the analysis of the available data trends, four different working regimes of the mill fan, called "cases" were defined [9] and shown in *table 2*.

In [10] the on-line measurable data were clustered using a "blind" separation algorithm from [11]. It also revealed four distinguishable working regimes of the mill fan. The results from "blind" clustering of the data were further used to derive a Takagi-Sugeno fuzzy rule base with a linear function of the input variables. The fuzzy rules have the following form:

If T is lt_i and V is lv_i and A is la_i Then case is C_i

Here lt_i, lv_i and la_i are linguistic values of the corresponding linguistic variables *T*, *V* and *A* in the *i*-th fuzzy rule, C_i is a linear combination of the crisp values of the input variables as follows:

$$C_i = a_i T + b_i V + d_i A + e_i$$

Here a_i, b_i, d_i and e_i are constants corresponding to *i*-th fuzzy rule. The fussy rule base was trained with ANFIS algorithm in Matlab using "blind" data classification [10] to

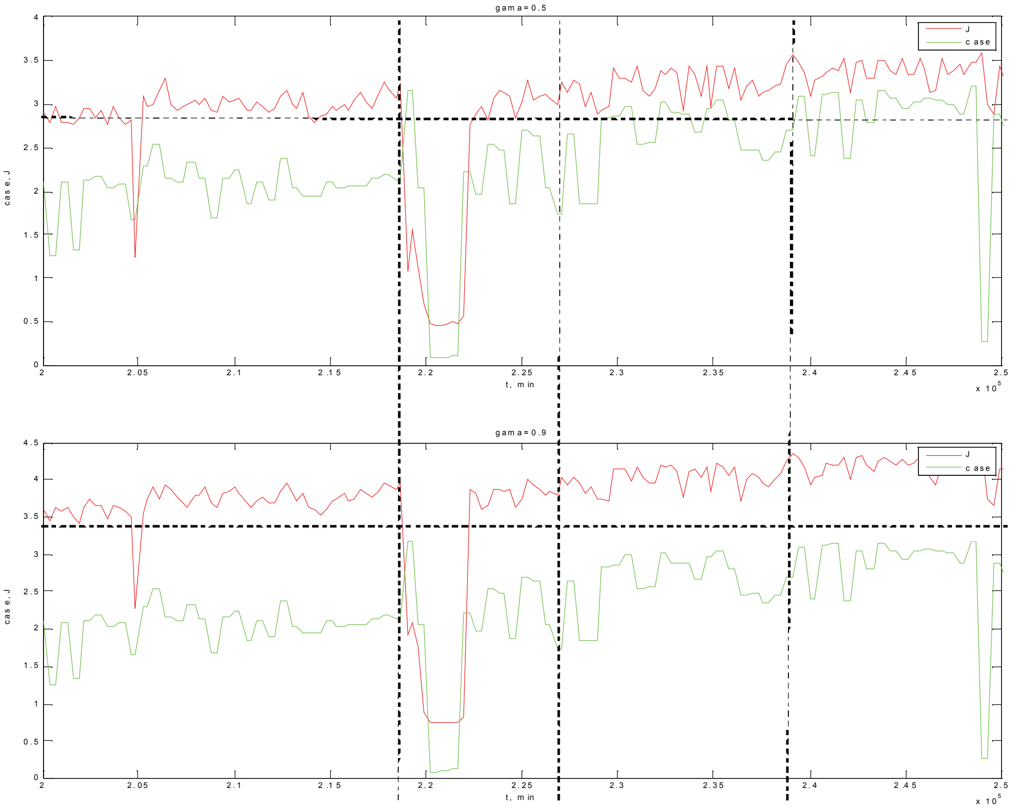

Figure 8. Predictions of the ESN critic in comparison to the real case for both γ=0.5 and γ=0.9

replace experts' knowledge. *Figure 4* represents the working regimes classification of all available data from the mill fan.

The approach proposed in the previous section for training of NN critic for predictive maintenance of the mill fan was tested. *Figure 5* represents the scheme for training of the critic for predictive maintenance. The vector of the plant state variables contains the available real time measurements, i.e., $S(k) = [V(k), T(k)]$. Hence, the overall input of the critic is $in(k) = [S(k)\ A(k)]$. The utility function was replaced by the case number $C(k)$. The fuzzy rule base was used to replace the plant operator that is intended to provide expert evaluation of the current working regime $C(k)$.

On-line RLS training of the critic ESN was applied since the previous investigations [7,8] showed better results in comparison to the batch algorithm. The training data is the same as the presented in *figure 4*.

Figure 6 and *figure 7* demonstrate predictions of the trained ESN critic in comparison to the real cases, as they are determined by the developed fuzzy classifier for two values of the discount factor – $\gamma = 0.5$ and $\gamma = 0.9$ respectively.

In order to demonstrate the preventive prediction of the critic network, "zoomed" parts of both *figure 6* and *figure 7* are compared in *figure 8*. The vertical dashed lines mark some points, from which it is clear that the critic output increases by several time steps, while the plant working regime number increases in both cases ($\gamma = 0.5$ and $\gamma = 0.9$). The horizontal dotted lines mark a value around which the critic output is found in both cases. It becomes clear that the bigger γ is, the bigger the output of the critic is, since the horizon of the discounted sum is bigger. However, there is no significant effect of the value of the discount factor on the time, when the prediction from the critic arrives.

Conclusion

The proposed here approach of ACD for predictive maintenance of an industrial plant allows training of a predictor of the possible alarm situations in the plant using only the available on-line measurements. It is not necessary to obtain an adequate and accurate plant model in order to predict correctly the future plant state, but it rather relies on a "learning by experience" approach. Additionally, adopting a fast trainable recurrent NN structure – ESN, allows also the on-line re-adaptation of the predictive critic network following the new information and accounting for the human expert advice.

The future work will be focused on development of the second level of the ACD scheme – the action network. It could be in the form of a rule-base that will serve as an operator advisor. These levels have to be trained in order to generate proposals for possible actions preventing any approaching alarm situation.

Acknowledgment

This work has been supported by the National Science Fund of Bulgaria under Project No. DVU-10-0267/10 and by Project AComIn "Advanced Computing for Innovation", Grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

References

1. Barto, A. G., R. S. Sutton, C. W. Anderson. Neuronlike Adaptive Elements that Can Solve Difficult Learning Control Problems. – *IEEE Trans. on Systems, Man and Cybernetics*, 13 (5), 1983, 834-846.

2. Bellman, R. E. Dynamic Programming. Princeton, NJ. Princeton Univ. Press., 1957.

3. Bertsekas, D. P., J. N. Tsitsiklis. Neuro-Dymanic Programming. Athena Scientific, Belmont, MA, 1996.

4. Doukovska, L., P. Koprinkova-Hristova, S. Beloreshki. Analysis of Mill Fan System for Predictive Maintenance. Int. Conf. Automatics and Informatics'11, 3-7 Oct. 2011, Sofia, Bulgaria, B-331–B-334.

5. Jaeger, H. Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the "Echo State Network" Approach. GMD Report 159, German National Research Center for Information Technology, 2002.

6. Jaeger, H. Adaptive Nonlinear System Identification with Echo State Networks. Advances in Neural Information Processing Systems 15 (NIPS 2002), Cambridge, MA, MIT Press, 2003, 593-600.

7. Koprinkova-Hristova, P., G. Palm. Adaptive Critic Design with ESN Critic for Bioprocess Optimization. Lecture Notes in Computer Science, 6353, 2010, 438-447.

8. Koprinkova-Hristova, P., M. Oubbati, G. Palm. Adaptive Critic Design with Echo State Net-work. Proc. of 2010 IEEE Int. Conf. on Systems, Man and Cybernetics, Istanbul, Turkey, 10-13 Oct. 2010, 1010-1015.

9. Koprinkova-Hristova, P., L. Doukovska, S. Beloreshki. Fuzzy Classifier of Mill Fan System Working Regimes. XIX Int. Symposium Process Control of Power Plant Systems, Bankya, Bulgaria, 2011, 25-28.

10. Koprinkova-Hristova, P., L. Doukovska, P. Kostov. Working Regimes Classification for Predictive Maintenance of Mill Fan Systems. 2013 IEEE International Symposium on Innovations in Intelligent Systems and Applications, IEEE INISTA 2013, 19-21 June 2013, Albena, Bulgaria, DOI: 10.1109/INISTA.2013.6577632.

11. Koprinkova-Hristova, P., N. Tontchev. Echo State Networks for Multi-dimensional Data Clustering. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7552 LNCS (PART 1), 2012, 571-578.

12. Lenardis, G. G. A Retrospective on Adaptive Dynamic Programming for Control. Proc. of Int. Joint Conf. on Neural Networks, Atlanta, GA, USA, 14-19 June 2009, 1750-1757.

13. Lukosevicius, M., H. Jaeger. Reservoir Computing Approaches to Recurrent Neural Network Training. – *Computer Science Review*, 2009, 3, 2009, 127-149.

14. Prokhorov, D. V. Adaptive Critic Designs and Their Applications. Ph.D. Dissertation. Department of Electrical Engineering, Texas Tech. Univ., 1997.

15. Schrauwen, B., M. Wandermann, D. Verstraeten, J. J. Steil. Improving Reservoirs Using Intrinsic Plasticity. – *Neurocomputing*, 71, 2008, 1159-1171.

16. Si, J., Y.-T. Wang. On-line Learning Control by Association and Reinforcement. – *IEEE Trans. on Neural Networks*, 12 (2), 2001, 264-276.

17. Sutton, R. S. Learning to Predict by Methods of Temporal Differences. – *Machine Learning*, 3, 1988, 9-44.

Manuscript received on 5.12.2013

Assoc. Prof. ***Petia Koprinkova-Hristova*** *received MSc degree in Biotechnics from the Technical University – Sofia in 1989 and PhD degree on Process Automation from Bulgarian Academy of Sciences in 2001. Since 2003 she is Associate Professor in the Institute of Control and System Research and from January 2012 – in the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences. Her main research interests are in the field of Intelligent Control Systems using mainly fuzzy, neuro-fuzzy and neural network approaches. Currently she is a member of European Neural Network Society (ENNS) executive committee for 2011-2016 and member of the John Atanasoff Society of Automatics and Informatics in Bulgaria.*

Contacts:
Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad G. Bonchev St., bl. 25A, 1113 Sofia
e-mail: pkoprinkova@bas.bg