Zbornik radova 53. Konferencije za ETRAN, Vrnjačka Banja, 15-18. juna 2009.
Proc. 53rd ETRAN Conference, Vrnjačka Banja, June 15-18, 2009

ONE STEP AHEAD PREDICTION IN ELECTRONICS BASED ON LIMITED INFORMATION

Jelena Milojković, jelena@venus.elfak.ni.ac.yu, http://leda.elfak.ni.ac.yu
Vančo Litovski, vanco@elfak.ni.ac.yu, http://leda.elfak.ni.ac.yu

Abstract – *One step ahead prediction based on short time series is discussed. Short time series are characterized with no trend information, no randomness and lack of periodicity. That makes prediction based on them very difficult or even impossible. On the other side there is strong need for prediction based on limited amount of data in many areas of life and business. We here propose implementation of some architectures of artificial neural networks as a potential systematic solution of that problem as opposed to heuristics that are in use. Examples will be given related to verification of Moor's law that is respected for prediction in modern electronic production, and to prediction of quantities of obsolete computers.*

1. INTRODUCTION

In an inspired paper [1] Prof. Mendel' clames: "Prediction of short time series is a topical problem. Cases where the sample length N is too small for generating statistically reliable variants of prediction are encountered every so often. This form is characteristic of many applied problems of prediction in marketing, politology, investment planning, and other fields." Further he claims: "Statistical analysis suggests that in order to take carefully into account all components the prediction base period should contain several hundreds of units. For periods of several tens of units, satisfactory predictions can be constructed only for the time series representable as the sum of the trend, seasonal, and random components. What is more, these models must have a very limited number of parameters. Series made up by the sum of the trend and the random component sometimes may be predicted for even a smaller base period. Finally, for a prediction base period smaller than some calculated value $N_{\min}$, a more or less satisfactory prediction on the basis of observations is impossible at all, and additional data are required".

Among the fields not mentioned in [1], dealing with really small set of data or "prediction base period", we will comment here first the prediction in the semiconductor industry that is considered to be governed by the Moor's law [2]. Almost all quantities related to modern electronics are supposed to be subject of Moor's law which is stated, in fact, relatively vaguely. Namely, by some authors the quantities governed by Moor's law change by factor of two every 18 months, while others are putting all this in a time frame between 15 and 24 months. This is why another general statement for the Moor's law is in use claming that exponential change is observed. Additional problems related to prediction are encountered when consideration of new phenomena is introduced [3] meanning that departures of the Moor's law are expected in the future in some aspects of the development of modern electronics.

The environmental impact of electronics became an important issue nowadays [4]. As a matter of fact, the eco-design of electrical and electronic products is already a legislative matter [5,6]. Electronic waste (EW) is considered hazardous while, in the same time, in enormous quantities. Prediction in this area is of paramount importance for planning and installing equipments, plants, and facilities for recycling and end-of-life management of electronic products, while short term data are available only.

Having no systematic methodology for prediction based on short prediction base period, even in the areas that are supposed to be governed by the Moor's law, heuristics were developed for prediction. In the best known example "International technology roadmap for semiconductors" [7], to predict the values (transistor dimensions, number of transistors per chip, number of transistors per pin etc.) for future, a so called *Scaling Calculator* is developed and implemented to approximate the Moor's law. Alternatively, in [8] an approximative expression of the form $y=a\cdot e^{bx}$ is implemented in order to confirm that even the number of computers sold per year in a country (Romania) obey the Moor's law.

In a set of recent studies [9,10,11,12] dealing with the quantities of EW, attempts were made to make prediction based on hunches and rules of thumb. In fact, some presumptions were maid and predictions based on them published. Later, having missed the target, the presumptions were corrected, and so on.

Having all that in mind we undertook a project of developing an ANN based method that will be convenient for systematic implementation in stationary time series prediction with reduced set of data. Our first results were published in [13,14]. The main idea implemented was the following. If one wants to create neural network that may be used for forecasting one should enable this property during ANN's training. In addition, the ANN used has to have such a structure to accommodate to the training process for prediction. Following these considerations new forecasting architectures were developed.

The goal of this paper is to put the new methods into a broader context of implementation of ANNs for forecasting and to verify them on a set of examples with different origins.

The structure of the paper is as follows. After general definitions and statement of the problem we will give a short background related to ANNs application to forecasting. Then we will describe two solutions for possible applications of ANNs aimed to the same forecasting task. Finally short discussion of the results and consideration related to future work will be given.

2. PROBLEM FORMULATION AND SOLUTION

A time series is a number of observations that are taken consecutively in time. A time series that can be predicted precisely is called deterministic, while a time series that has future elements which can be partly determined using previous values, while the exact values cannot be predicted, is said to be stochastic. We are here addressing only deterministic type of time series.

Consider a scalar time series y^i, i=1,2, … m. It represents a set of observables of an unknown function $\hat{y} = \hat{f}(t)$, taken at equidistant time instants separated by the interval Δt i.e. $t^{i+1} = t^i + \Delta t$. One step ahead forecasting means to find such a function f that will perform the mapping

$$(1) \qquad y^{m+1} = f(t^{m+1}) = \hat{y}^{m+1} + \varepsilon ,$$

where $\hat{y}^{m+1}$ is the desired response, with an acceptable error ε.

The prediction of a time series is synonymous with modeling of the underlying physical process responsible for its generation. This is the reason of the difficulty of the task. There have been many attempts to find solution to the problem. Among the classical deterministic methods we may mention the k-nearest-neighbor [15], in which the data series is searched for situations similar to the current one each time a forecast needs to be made. This method asks for periodicity to function that is not the case in the situation considered in this paper.

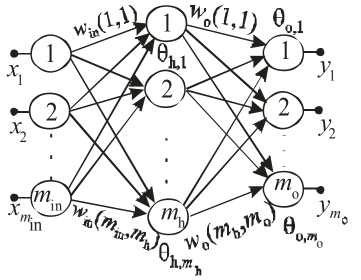

Fig. 1. *Fully connected feed-forward neural network with one hidden layer and multiple outputs*

In the past decades ANNs have emerged as a technology with a great promise for identifying and modeling data patterns that are not easily discernible by traditional methods. A comprehensive review of ANN use in forecasting may be found in [16]. Among the many successful implementations we may mention [17]. A common feature, however, of the existing application is that they ask for a relatively long time series to become effective. Typically it should be not shorter then 50 data points [16]. This is due to the fact that they all look for periodicity within the data. Very short time series were treated [17]. Here additional "nonsample information" was added to the time series in order to get statistical estimation from deterministic data.

That is why we went for a search for topological structures of ANN that promise prediction based on short time series. In the next, we will first briefly introduce the feed-forward neural networks that will be used as a basic structure for prediction throughout this paper.

The network is depicted in Fig. 1. It has only one hidden layer, which has been proven sufficient for this kind of problem [18]. Indices: "in", "h", and "o", in this figure, stand for input, hidden, and output, respectively. For the set of weights, $w(k, l)$, connecting the input and the hidden layer we have: $k=1,2,..., m_{in}$, $l=1,2,..., m_h$, while for the set connecting the hidden and output layer we have: $k=1,2, ...m_h$, $l=1,2,..., m_o$. The thresholds are here denoted as $\theta_{x,r}$, $r=1,2, \ldots, m_h$ or m_o, with x standing for "h" or "o", depending on the layer. The neurons in the input layer are simply distributing the signals, while those in the hidden layer are activated by a sigmoidal (logistic) function. Finally, the neurons in the output layer are activated by a linear function. The learning algorithm used for training is a version of the steepest-descent minimization algorithm. The number of hidden neurons, m_h, is of main concern. To get it we applied a procedure that is based on proceedings given in [19].

In prediction of time series, in our case, a set of observables is given (per year) meaning that only one input signal is available, the discretized time. According to (1) we are predicting one quantity at a time meaning one output is needed, too. The values of the output are numbers (transistor size, memory capacity, millions of pieces or weight of obsolete computer units). To make the forecasting problem numerically feasible we performed transformation in both the time variable and the response. The time was reduced by t_0-1 so that

(2) $t=t^*-(t_0-1)$.

Having in mind that t^* stands for the year, this reduction gives the value of 1 to the year (t_0) related to the first sample. The samples are normalized in the following way

(3) $y=y^* / M$

where y^* stands for the current value of the target function, M is a constant (for example $M=10^6$ cubic feet for the amount of obsolete computers).

If the architecture depicted in Fig. 1 was to be implemented the following series would be learned: $(t^i, f(t^i))$, $i=1,...,m$.

Starting with the architecture of Fig. 1, in [14] possible solutions were investigated and two new architectures were suggested to be the most convenient for the solution of the forecasting problem based on short prediction base period.

The first one, named *time controlled reccurent* (TCR) was inspired by the time delayed recurrent ANN. It is a recurrent and time delayed architecture but, in the same time, insists on the time variable to control the predicted value as depicted in Fig. 2. Our intention was to benefit from both: the generalization property of the ANNs and the success of the recurrent architecture. Here in fact, the network is learning a set in which the output value is controlled by the present time and its own previous instances:

(5) $y^{i+1} = f(t^i, y^i, y^{i-1}, y^{i-2})$, $i=3, ...,m$.

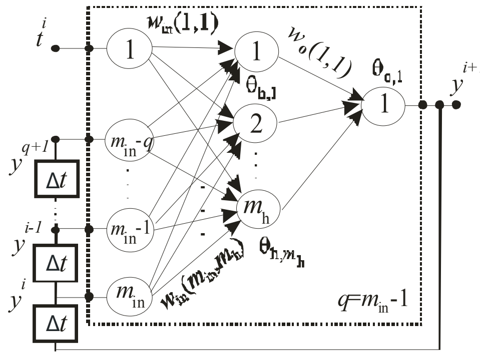

Fig. 2. *TCR. Time controlled recurrent ANN*

The second architecture is named *feed forward accommodated for prediction* (FFAP) and depicted in Fig. 3. Our idea was here to force the neural network to learn the same mapping several times simultaneously but shifted in time. In that way, we suppose, the previous responses of the function will have larger influence on the $f(t)$ mapping.

The architecture is depicted in Fig. 3. There is one input terminal that, in our case, is t^i. The *Output*$_3$ terminal, or the future terminal, in our case, is to be forced to approximate y^{i+1}. In cases where multiple-step prediction is planned *Output*$_3$ may be seen as a vector. *Output*$_2$ should represents the present value i.e. y^i. Finally, *Output*$_1$ should learn the past value i.e. y^{i-1}. Again, if one wants to control the mapping by a *set* of previous values, *Output*$_1$ may be seen as a vector.

As an example we may express the functionality of the network as

(6) $\{y^{i+1}, y^{i}, y^{i-1}, y^{i-2}\} = \mathbf{f}(t^{i})$, i=3, ...,m,

where $Output_1$={ y^{i-1}, y^{i-2} }, meaning that: one future, one present and two previous responses are to be learned.

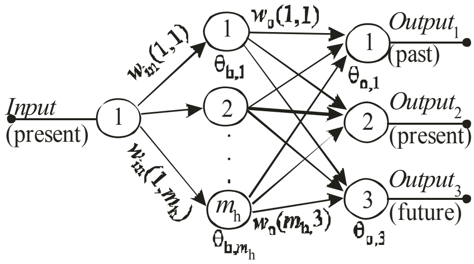

Fig. 3. *FFAP. Feed forward ANN structure accommodated for prediction*

3. IMPLEMENTATION EXAMPLES

Three examples will be given here demonstrating the properties of the solutions proposed.

First, we will consider the prediction of the quantities of obsolete computers in the USA based on data given in [12]. According to [12], putting t_0= 1991, after normalization, we get Table 1 as the set of observables representing the quantities of obsolete computers Hin the USA. The same data are visualized in Fig. 4. Here M=10^6 cubic feet.

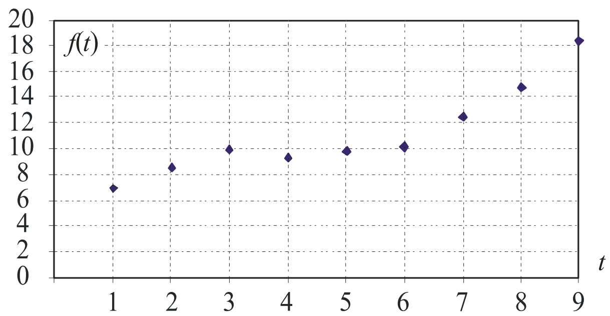

Fig. 4. *Visualization of Table 1*

Table 1. *Quantities of obsolete computers in time*

t	1	2	3	4	5	6
f(t)	7.03	8.67	10.0	9.33	9.85	10.18
t	7	8	9			
f(t)	12.54	14.76	18.4			

The first eight will be used as training data while the last one i.e. t=9 and $f(t)$=18.4, will be compared with the predictions obtained, in order to validate the methods.

In the following, two experiments will be described based ANN architectures emanated from Fig. 2 and Fig. 3.

The results obtained after learning are expressed in Table 2. It contains information on both the structure of the networks and the values obtained by prediction.

Table 2. *Prediction of quantities of obsolete computers. Note:* $\hat{f}(9) = 18.4$.

Solution type	No. of hidden neurons	No. of output neurons	$f(9)$
TCR	10	1	17.2114
FFAP	4	4	18.2274

By examining the results depicted in Table 2 we may conclude that satisfactory prediction was obtained with both architectures. Nevertheless, it is to be mentioned that the FFAP is considerably nearer to the solution needed. What is not expressed in the table is the fact that the FFAP solution is much more sensitive to the initial solution for the weights and thresholds, making the training process more difficult and uncertain.

It is not shown here, for the sake of simplicity, but it is worth mentioning that both TCR and FFAP approximate excellent. That means that except for the [9, $f(9)$] point, all previous points on the curve $f(t)$ overlap exactly with the ones depicted in Fig. 4.

As a second example the data published in [7] for the physical gate length (L) trends will be used. In Table 3, ten values of L are given for a ten year period starting with 1996. The value for the year 2005, $\hat{f}(10) = 36$, was to be matched by prediction. Again, two architectures were used the results being shown in the same table and visualized in Fig. 5. As can be seen the TCR and FFAP architectures produce almost the same prediction while getting excellent approximation.

Table 3. *MOS transistor's physical gate length (in a microprocessor) . M=10^{-9} m.*

t	Value to be matched	TCR	FFAP
1	265		
2	186.56		
3	127.1	127.1	127.099
4	90.85	90.8499	90.8485
5	74.98	74.9799	74.979
6	63.4	63.4	63.3996
7	53.6	53.6	53.5997
8	44.24	44.24	44.2397
9	37.41	37.41	37.4097
10	31.6	32.571	32.6725

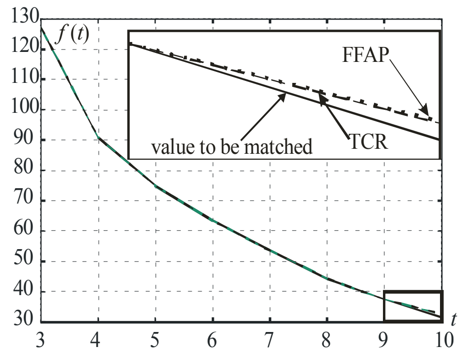

Fig. 5. *Visualization of Table 3*

Table 4. *Number of MOS transistors per chip (in a microprocessor). M=10^6.*

t	value to be matched	TCR	FFAP
1	12.3		
2	16.8		
3	24.3	24.3004	24.3
5	48.9	48.9004	48.9
7	96.1	96.1005	96.0999
8	137.3	137.3	137.1
9	192.5	169.42	193.133

As the last example here we will present the results obtained for the prediction of the number of transistors per chip in a microprocessor IC. These numbers are to be considered as the generic ones expressing the Moor's law.

Following the procedure as above and starting again with the year 1996, we got the results depicted in Table 4 (note that the years 4 and 6 are skiped). All these are visualized in Fig. 6.

While, again, excellent approximation is obtained with both networks, one may notice that FFAP predicts much better in this case which may be characterized as one processing extremely reduced base period data.

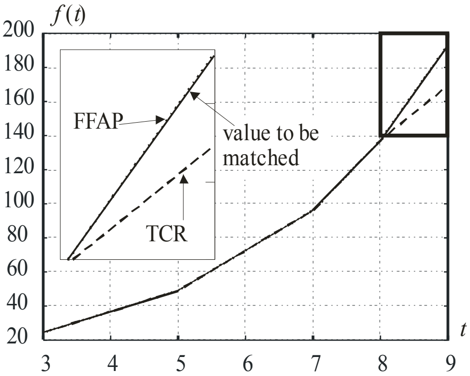

Fig. 6. *Visualization of Table 4*

4 CONCLUSION

Two new ANN architectures were proposed here for the solution of the forecasting problem in cases with reduced base period and found to be promising for further research. The FFAP architecture seems to get better results but it is experienced that it is more difficult to be generated in terms of convergence of the training process and jumping into local inadequate solutions. It is our experience that one should implement the following procedure in order to get the best solution. One is first to get the prediction by the TCR method and use it as a reference. Then one is to choose among the solutions obtained by FFAP method the one nearest to the TCR solution. As a future investigation we are considering averaging the FFAP solutions being clustered around the TCR solution, if such cluster is obtained. We also consider the possibility to use the result of the one step prediction method for multistep prediction by pronouncing it an input data for the next time instant.

5. REFERENCES

[1] A. S. Mandel', "Method of Analogs in Prediction of Short Time Series: An Expert-statistical Approach", *Automation and Remote Control*, Vol. 65, No. 4, April 2004, 634-641.

[2] G. E. Moore, "Cramming more components onto integrated circuits", *Electronics*, Vol. 38, No. 8, April 19, 1965

[3] J.D. Meindl, "Beyond Moore's Law: the interconnect era**"**, *Computing in Science & Engineering***,** ISSN: 1521-9615, Vol. 5, No. 1, Jan./Feb. 2003, pp. 20- 24.

[4] B. Karn, H. S.Matthews, "Nanotechnology: Emerging Challenges for Electronics and the Environment", *IEEE Spectrum*, Vol. 44, No. 9, 2007, pp. 54-58.

[5] -, Directive 2005/32/ec of the European parliament and of the council of 6 July 2005, 22.7. 2005, *EN Official Journal of the European Union L* 191/29.

[6] A. Bandyopadhyay, "A regulatory approach for e-waste management: a cross-national review of current practice and policy with an assessment and policy recommendation for the Indian perspective", *International Journal of Environment and Waste Management*, Volume 2, No. 1-2, 2008, pp. 139 – 186.

[7] -, " International technology roadmap for semiconductors, 2007 edition, Executive summary", http://www.itrs.net/.

[8] Mateescu, G.D., et all. "A Short-Time Prediction of The Romanian Personal Computers Market Based On The Moore Law", *Journal for Economic Forecasting*, Sept. 2003, pp. 93-97.

[9] V. B. Litovski, et all., "Program of establishment of the recycling system of waste electronic of computers", *Serbian Agency for Recycling*, Belgrade, 2006, (in Serbian).

[10] H. S. Matthews and D. Matthews, "Computers and the Environment: Understanding and Managing Their Impacts", in: E. Williams and R. Kuehr, eds., *Computers and the Environment*, Kluwer, Dordrecht, 2004.

[11] K. Feszty, M. Colin, and J. Baird, "Assessment of the quantities of waste electrical and electronic equipment (WEEE) in Scotland", *Waste management & research,* ISSN 0734-242X, 2003, vol.21, No. 3, pp.207-217.

[12]H. S.,Matthews, et all., "Disposition and End-of-Life Options for Personal Computers", *Green Design Initiative Technical Rrt. #97-10*, Carnegie Mellon University, 1997.

[13] J. Milojković, V. B. Litovski, "New methods of prediction implemented for sustainable development", *51th Conf. ETRAN*, Herceg Novi, 2007, Paper no. EL1.8 (in Serbian).

[14] J. Milojković, V. B. Litovski, "Comparison of some ANN based forecasting methods implemented on short time series", *9th Symposium NEUREL-2008*, Belgrade, ISBN 978-1-4244-2903-5, 2008, pp. 175-178.

[15] E.A. Plummer, "Time series forecasting with feedforward neural networks: guidelines and limitations", M.S. Thesis, University of Wyoming, Laramie, July 2000.

[16] Zhang, B.G., "Forecasting with artificial neural networks: The state of the art", *International Journal of Forecasting*, Vol. 14, No. 1, March 1998, pp. 35-62

[17] K. Brännäs and J. Hellström, "Forecasting based on Very Small Samples and Additional Non-Sample Information", Umeå University, Sweden, 1998

[18] T. Masters, "Practical Neural Network Recipes in C++", *Academic Press,* San Diego, 1993.

[19] E.B. Baum and D. Haussler, "What size net gives valid generalization", *Neural Computing*, 1989, Vol. 1, pp. 151-160.

PREDVIĐANJE U ELEKTRONICI JEDAN KORAK UNAPRED ZASNOVANO NA OGRANIČENOM PREDZNANJU

Jelena Milojković i Vančo Litovski

Sažetak: Razmatrano je predviđanje jedan korak unapred zasnovano na kratkom vremenskom nizu. Kratki vremenski nizovi karakterišu se nedostatkom informacije o trendu, nedostakom periodičnosti i nedostakom stohastičke komponente. To čini predviđanja zasnovana na njima veoma teškim ili nemogućim. S druge strane, u mnogim područjima kao što je marketing, politologija, planiranje investicija, planiranje ekološkog razvoja, planiranje tehnološkog razvoja, postoji značajna potreba za predviđanjem zasnovanim na ograničenoj količini podataka. Ovde predlažemo dve arhitekture veštačkih neuronskih mreža kao moguća rešenja ovog problema. Dati su primeri koji se odnose na verifikaciju Moorovog zakona kao i na predviđanje količina zastarelih računara. Pokazano je da je upravo nedostatak heurističke komponente taj koji omogućava uspešno predviđanje.