Elman Recurrent Neural Network in Thermal Modeling of Power Transformers

M. HELL, F. GOMIDE
Department of Computer Engineering and Industrial Automation - DCA
School of Electrical and Computer Engineering - FEEC
State University of Campinas - UNICAMP
Av. Albert Einstein, 400, Cidade Universitria Zeferino Vaz, 13083-970, Campinas, SP
BRAZIL

L. SECCO, and P. COSTA Jr.
Graduate Program in Electrical Engineering
Pontifical Catholic University of Minas Gerais
Av. Dom José Gaspar, 500, Prédio 3, 30535-610, Belo Horizonte, MG
BRAZIL

Abstract: — This work suggests a Elman Recurrent Networks (ELRN) as a means to model thermal condition of power transformers. Experimental results with actual data reported in the literature show that ELRN modeling requires less computational effort, and is more robust and efficient than multilayer feedforward networks, radial basis function network, and classic deterministic modeling approaches.

Key-Words: — Power transformers; recurrent networks; Elman networks; thermal modeling.

1 Introduction

Power transformers are key pieces in transmission and distribution of electric energy. Transformer failures can profoundly affect power systems, causing potentially widespread power outages and, consequently, economic and social losses. Condition monitoring is very important to guarantee safe operation and to increase power availability and service life of power transformers.

One main factor in power transformer operation concerns its working temperature, that is, its *hot-spot temperature*, the temperature at the top or in the center of the high or low voltage windings. The hot-spot temperature has a key role in insulation aging and in service life of power transformers. Monitoring of hot-spot temperature is vital to develop overload protection systems. High hot-spot temperature means acceleration in transformer aging.

Dynamic modeling of the thermal behavior is one of the most important aspects to monitor power transformer conditions. Comparisons between measured and predicted working temperature provide information about transformer conditions and indicate potential anomalies.

Load capability of power transformers can be found using transient heat equations and their specific thermal characteristics [4]. Currently, the use of conventional modeling techniques to determine the internal temperature of power transformers is a challenging task. Security induces adoption of conservative safety factors in computation and often this means underutilized transformers once hot-spot portions of the coils are kept from overheating and failing prematurely. Consequently, the maximum power transfer computed may be 20-30% less than nominal power transformer capacity [5].

To increase system operational margins during overload periods, new loading approaches are needed to bring power transformers beyond the corresponding nominal specifications. The approaches should be capable to timely change load capability rates to use full transformer capacity, meaning less revenue losses and lower maintenance and upgrade costs.

Several modeling techniques are available today, artificial neural networks-based modeling (ANN) being among the most relevant and efficient. ANN models, especially multilayer feedforward structures, have been used extensively in system modeling due to its ability to learn complex non-linear relationships that are difficult for conventional approaches [8].

Feedforward neural models do not have internal dynamics. They are static, may not comply with actual system dynamics and fail to provide efficient models. Despite the efficiency of supervised learning schemes, the ones that use input/output system data, imprecision and inadequacy of input/output data [8] may potentially harm feedforward ANN performance.

An alternative that has received considerable attention and shown significant progress during the last years concerns recurrent network modeling approaches. Similar to static feedforward neural networks, dynamic recurrent neural neural networks consist of global recursive network and partial recursive network [1]. In the global recursive network, there are full connections between each two nodes, and all weights are trainable. Because complex recurrent connections of the nodes in fully recurrent neural networks result in poor convergence speed. In contranst, in the partial recursive network, feedforward loop is the basic structure in which the weights are variable, and the back-forward loop consists of "Context" in which the weights are fixed [2]. So it is investigated widely and deeply.

Elman neural network is a partial recurrent network model introduced by Elman [3] that lies somewhere between a classic feddforward multilayer perceptron and a pure recurrent network. Because of the existence of contest nodes and local recurrent connections between the context layer and the hidden layer, it has certain dynamical advantages over static neural networks, such a Multi-Layer Perceptron (MLP) and Radial Basis Function Networks (RBFN). This highlight makes Elman neural network very suitable to be utilized in the dynamic system modeling problem.

This paper suggests a Elman Recurrent Network (ELRN) as an alternative to model the thermal condition of power transformers. Experimental results reported here show that the Elman model is more effective than multilayer feedforward backpropagation network (MLP), radial basis function network (RBF), and a classic deterministic model. Using actual data reported in the literature, the ELRN network provides an effective and robust model, learns very quickly, and efficiently manages learning data imprecision.

2 Deterministic Modeling

Nowadays, as mentioned above, load capability is found using transient heating equations and specific thermal characteristics and parameters of power transformers. Load capability calculation requires knowledge of load curves and operating conditions (especially the operating temperature), whose values are fixed and usually conservative [5].

As discussed in [4], load capability calculations require the following variables and parameters:

Variables (functions of time, t):

Θ_A = environment temperature, $°C$.
Θ_{TO} = top oil temperature, $°C$.
Θ_H = hot-spot temperature, $°C$.
$\Delta\Theta_H$ = hot-spot rise above top oil temperature, $°C$.
$\Delta\Theta_{H,U}$ = ultimate hot-spot temperature rise over top oil (for a given load current), $°C$
$\Delta\Theta_{TO,U}$ = ultimate top oil temperature rise over environment (for a given load current), $°C$
K = load current, per unit.

Parameters (constants):

$\Delta\Theta_{TO,R}$ = rated top oil temperature rise over environment, $°C$
$\Delta\Theta_{H,R}$ = rated hot-spot temperature rise over top oil, $°C$
τ_{TO} = top oil rise time constant, hours
τ_H = hot-spot rise time constant, hours
R = ratio of load loss at rated-load to no-load loss at applicable tap position, dimensionless
m = empirically derived value, depends on the cooling method, dimensionless.
n = empirically derived value, depends on the cooling method, dimensionless.

Using these variables and parameters, the heat transfer equations and the step-by-step load capability calculation process are as follows:

- At each time step, compute the *ultimate top oil rise*($\Delta\Theta_{TO,U}$) using the *load current* value at that instant and:

$$\Delta\Theta_{TO,U} = \Delta\Theta_{TO,R}\left[\frac{K^2R+1}{R+1}\right]^n \quad (1)$$

- From (1), and the *environment temperature* at each time step, compute the increment in the *top oil temperature*, using the differential equation:

$$\tau_{TO}\frac{d\Theta_{TO}}{dt} = [\Delta\Theta_{TO,U} + \Theta_A] - \Theta_{TO} \quad (2)$$

in its finite difference equation form:

$$D\Theta_{TO} = \frac{Dt}{\tau_{TO}}\left([\Delta\Theta_{TO,U} + \Theta_A] - \Theta_{TO}\right) \quad (3)$$

where the prefix D implies a small finite step.

- Next compute the *ultimate hot-spot rise* using:

$$\Delta\Theta_{H,U} = \Delta\Theta_{H,R}K^{2m} \quad (4)$$

- The increment in the *hot-spot temperature rise* is found using the differential equation:

$$\tau_H\frac{d\Delta\Theta_H}{dt} = \Delta\Theta_{H,U} - \Delta\Theta_H \quad (5)$$

in its finite difference equation form:

$$D\Delta\Theta_H = \frac{Dt}{\tau_H}[\Delta\Theta_{H,U} - \Delta\Theta_H] \quad (6)$$

- Finally, add the *top oil temperature* to the *hot-spot rise* to get the *hot-spot temperature*, that is, set:

$$\Theta_H = \Theta_{TO} + \Delta\Theta_H \quad (7)$$

The model described by (1) - (7), presumes some simplifying assumptions such as: the oil temperature profile inside winding increases linearly from bottom to top; the difference between the winding temperature and the oil temperature is constant along the winding; the hot-spot temperature rise is higher than the temperature rise of the wire at the top of the winding, introducing a conservative factor; the environment temperature drives the oil temperature up and down with the same time constant as does the winding temperature; the incidence of solar flux is neglected.

Therefore, the use of the deterministic thermal model (1) - (7) may produce substantial error when computing load capability rate in real-time. Thus, there is a need to adopt new, more precise and robust methods to compute load capability and use it as a means to increase system operation margin in presence of overload conditions. In what follows, we suggest an alternative based on a Elman recurrent network.

3 Elman Network Structure

The basic structure of Elman neural network is composed by four layers: input layer, hidden layer, context layer, and output layer, as depicted in the figure 1. There are adjustable weights connecting each two neighboring layers. Generally, it is considered as a special kind of feedforward neural network with additional memory neurons and local feedback [6].

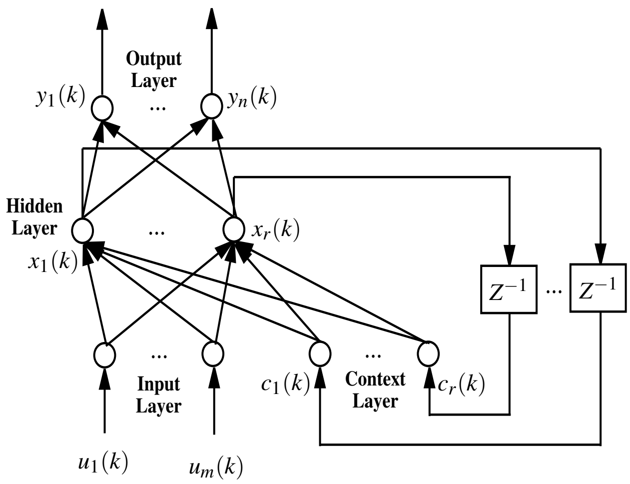

Figure 1. Structure of the Elman network

The self connections of the context nodes in the Elman network make it also sensitive to the history of input data which is very useful in dynamic modeling [7].

The notation used in this section is given below:

$W1_{ij}$	=	The weight that connects node i in the input layer to node j in the hidden layer
$W2_{jq}$	=	The weight that connects node j in the hidden layer to node q in the output layer
$W3_{lj}$	=	The weight that connects context node l to node j in the hidden layer
m,n,r	=	The number of nodes in the input, output and hidden layers respectively.
$u_i(k), y_j(k)$	=	Inputs and outputs of the Elman network, where $i = 1,\cdots,m$ and $j = 1,\cdots,n$
$x_i(k)$	=	Output of the hidden node i, $i = 1,\cdots,r$
$c_i(k)$	=	Output of the context node i, i.e., the output of the hidden node i of last time.
Z^{-1}	=	A unit delay.

For each unit in the hidden layer an additional unit called context unit is added. The context unit is fully connected with all the hidden units in a forward manner. This means that there is a weight from every context unit to every hidden unit. Furthermore, there are recurrent connections from the hidden units back to the context units. But each hidden unit is only connected to its associated context unit.

The weights of the recurrent connections are fixed and the forward weights get trained by using backpropagation. In the forward phase the context units behave like input units. The values of the hidden units and of the output units get calculated in the same way as it is done for feedforward networks. After calculating the outputs of the hidden units, the current values get copied into the correspondent context unit via recurrent connections (through a unit delay). These values are used in the next time step. At the first time step they have to be set to some initial values. During the backward phase of the training, target values for the outputs are used and the forward weights are adjusted by backpropagation.

The inputs of network are: $u(k) \in \mathbb{R}^m$, $y(k) \in \mathbb{R}^n$, $x(k) \in \mathbb{R}^r$, the outputs in each layer can be given by:

$$\begin{aligned} x_j(k) &= f\left(\sum_{i=1}^{m} w1_{ij}u_i(k) + \sum_{l=1}^{r} w3_{lj}c_l(k)\right) \\ c_l(k) &= x_j(k-1) \\ y_q(k) &= g\left(\sum_{j=1}^{r} w2_{jq}x_j(k)\right) \end{aligned} \tag{8}$$

where, $f(\bullet)$ and $g(\bullet)$ are the linear or nonlinear output function of hidden layer and output layer respectively.

Because the dynamic characteristics of Elman network are provided only by internal connection, so it needn't use the state as input pr training signal. This is the advantage of the Elman network in contrast with static feedforward network.

4 Simulation Results

In this section the recurrent Elman model presented in the previous section is used to estimate the hot-spot temperature of an actual transformer. The data set adopted during the experiments reported in this paper is the same of [5]. The data were collected from measurements performed in an experimental power transformer whose characteristics are summarized in Table 1.

The results are compared with three alternative models: the deterministic model (DM) summarized in Section 2, a backpropagation feedforward multilayer neural network (MLP), and a radial basis function neural network (RBFN) as used in [5].

All four models (deterministic, backpropagation, radial basis and Elman) were trained using data describing the behavior of transformer hot-spot temperature during a horizon of 24 hours with 5 minutes sampling period. The learning data is depicted in Figure 2.

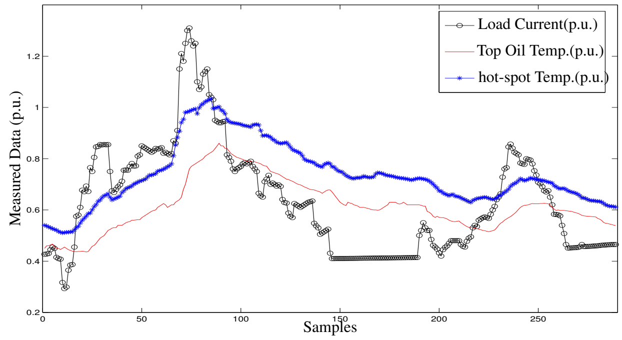

Figure 2. Learning data

After learning, the approximation, generalization and robustness capabilities of the networks were tested using two data sets describing different load capability conditions than those in learning data [5]. Test date, shown in Figure 3, were collected in the same way as training data.

Figures 4 and 5 depict the results provided by the deterministic (DM), Elman recurrent network (ELRN), backpropagation multilayer network (MLP) and the radial basis function network (RBFN) models. Figure 4 shows the actual transformer output and the corresponding models outputs under no overload condition whereas 5 shows the same outputs, but under overload conditions.

Table 2 summarizes the performance index values achieved by the models during the simulation experiments. In the table, $MSE-Dt1$ is the mean square error (MSE) obtained when running the models using the test data with no overload condition (Figure 3(a)), $MSE-Dt2$ is the MSE obtained when using

Table 1. Main Characteristics of the Power Transformer

Nameplate rating	25 kVA
Vprimary / Vsecundary	10 kV / 380 V
Iron losses	195 W
Copper losses (full load)	776 W
Top oil temp. rise at full load	73.1 $^{\circ}C$
Length x width x height of the tank	64 x 16 x 80 cm
Type of cooling	ONAN*
Factory / year	MACE/87

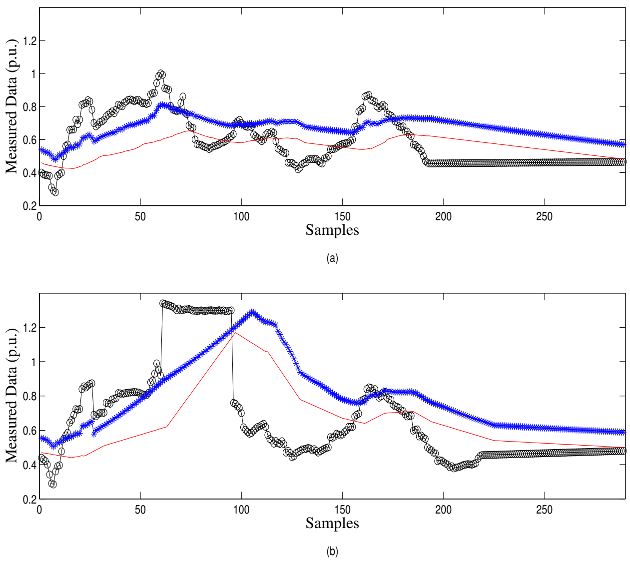

Figure 3. Test data: (a) no overload (b) with overload

the models with test data with overload condition (Figure 3(b)), and $MSE-Dts$ is the MSE obtained when running the models using both data sets.

Table 2. Simulation Results

Model	Learning Time (seg.)	MSE-Dt1	MSE-Dt2	MSE-Dts
MD	-	$17,3489$	$6,7822$	$12,0655$
ELRN	59,39	$0,4780$	$0,2,1073$	$1,2926$
MLP	92,76	$0,7901$	$2,4885$	$1,6393$
RBFN	82,83	$0,2565$	$0,9917$	$0,6241$

As Figures 4 and 5 show, ELRN, MLP and RBFN models provide good results when modeling the hot-spot temperature. However, MLP and RBFN require complex learning processes whereas the recurrent model (ELRN) uses a faster and simpler learning procedure. On the other hand, when training data do not assemble an ideal learning data set, a data set that it is not representative of the target behavior, the ELRN outperforms MLP and RBFN because of its robustness when dealing with imprecision in data.

To verify this important issue, we use a data set that does not fully represent the transformer dynamics. Next, we train the MLP, ELRN and RBFN neural networks using this same data set. In our experiment, as an extreme case, we took the data shown in Figure 3(a) to train the models and data of Figure 3(b) to verify the approximation, robustness and generalization capabilities. Figure 6 shows the results provided by the models in this circumstance. Table 3 summarizes

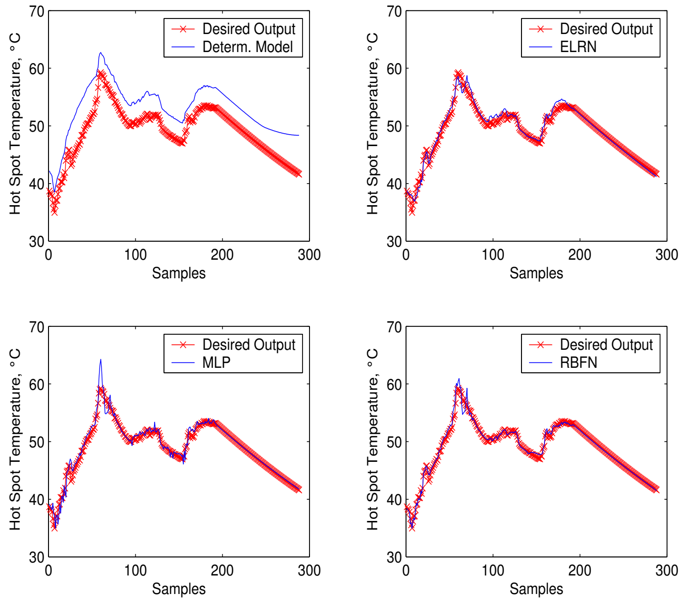

Figure 4. Transformer and models outputs with no overload condition

the mean square errors in this case. As we can easily see, the behavior of both, the MLP and ELRN models have shown substantial changes. Contrary, the recurrent model was able to keep the same hot-spot temperature prediction performance as before. Clearly, the recurrent model approximates the true transformer behavior closely, is able to generalize properly, and is more robust to data imprecision than its counterparts.

Table 3. Results using non ideal training data

	MD	ELRN	MLP	RBFN
MSE-Dt3	$6,7822$	1.8896	$1,2487\times10^2$	$4,6579\times10^4$

5 Conclusion

This paper has introduced an alternative Elman recurrent network to model the thermal condition of power transformers. Simulation experiments have shown that the Elman model is more effective than multilayer feedforward backpropagation network (MLP), radial basis function network (RBF), and a deterministic model because it approximates and generalizes transformer dynamics properly and is able to manage imprecise data. The recurrent network provides an effective and robust model, learns quickly, and requires modest computational power. Therefore, it is a candidate to safely contribute to increase power transformer real-time load capability during overload periods. In

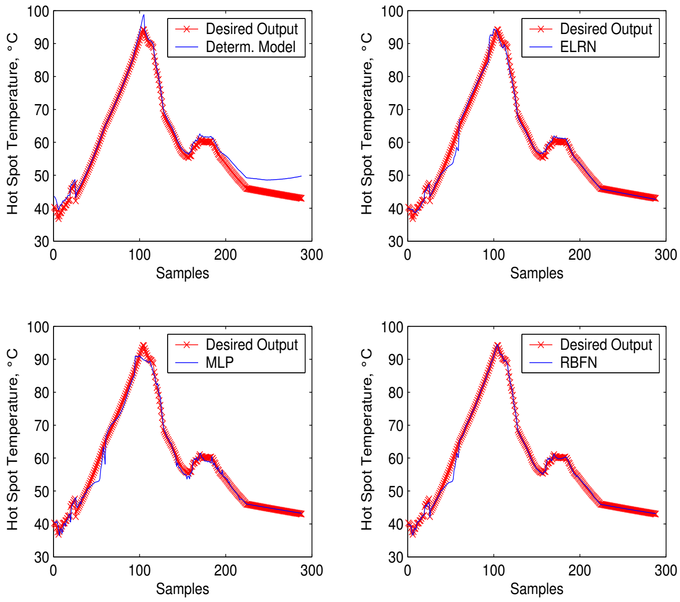

Figure 5. Transformer and models outputs with overload condition

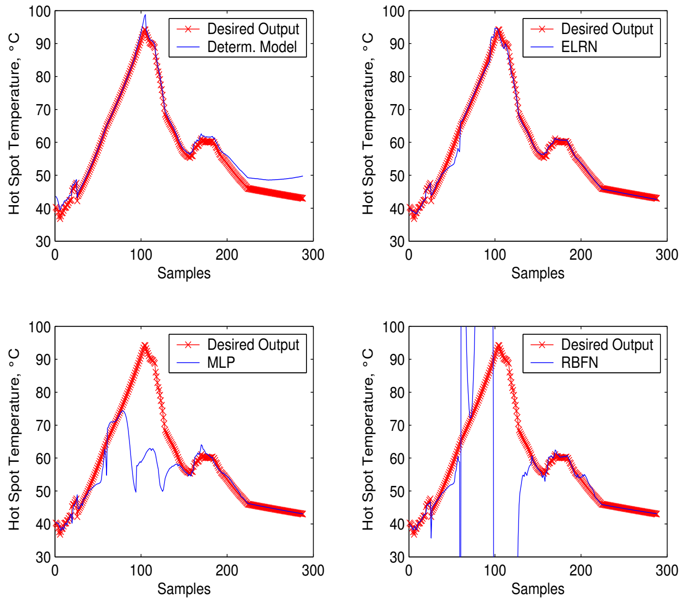

Figure 6. Models robustness evaluation

the future we expect to use a similar approach to estimate the loss in service lifes of transformers when submitted to overload conditions.

Acknowledgment

The authors acknowledge FAPESP, the Research Foundation of the State of São Paulo, for fellowship 03/05042-1, CNPq, the Brazilian National Research Council for grant 300729/86-3, the Pontifical Catholic University of Minas Gerais (IPUC, PUC-MG) for its support, and Álvaro Martins and Ronald Moura, engineers of the Energetic Company of Minas Gerais (CEMIG), for their comments during the development of this work.

References:

[1] C.Lee Giles, Gary M.Kuhn, Ronald J.Willians, *Dynamic Recurrent Neural Networks: Theory and Applications*, IEEE Transactions on Neural Networks, Vol. 5, No. 2, pp. 153-155, 1994.

[2] Y. Cheng, W. Qi, W. Cai, *Dynamic Properties of Elman and Modified Elman Neural Networks*, Proceedings of the First International Conference on Machine Learning and Cybernetics, Bejing,pp. 637-640, Nov. 2002.

[3] J. Elman, *Finding Structure in Time*, Cognitive Science, Vol. 14, pp. 179-211, 1990.

[4] G. W. Swift et. al., *Adaptative Transformer Thermal Overload Protection*, IEEE Transactions on Power Delivery, Vol. 16, No. 4, pp. 516-521, 2001.

[5] V. Galdi, L. Ippolito, A. Piccolo and A. Vaccaro, *Neural Diagnostic System for Transformer Thermal Overload Protection*, IEE Proceedings of Electric Power Applications, Vol. 147, No. 5, pp. 415-421, 2000.

[6] P.S. Sastry, G. Santharam, K.P. Unnikrishnan, *Memory Neuron Networks for Identification and Control of Dynamic Systems*, IEEE Transactions on Neural Networks, Vol. 5, pp. 306-319, 1994.

[7] K.S. Narendra, K. Parthasarathy, *Identification and Control of Dynamic Systems using Neural Networks*, IEEE Transactions on Neural Networks, Vol. 1, pp. 4-27, 1990.

[8] G. Goodwin and R. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, Inc., New York-USA, 1977.

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, NJ-USA, ed. 2, 1998.