Importance of input data normalization for the application of neural networks to complex industrial problems

Article *in* IEEE Transactions on Nuclear Science · July 1997
DOI: 10.1109/23.589532 · Source: IEEE Xplore

Joaquin Sevilla
Universidad Pública de Navarra

IMPORTANCE OF INPUT DATA NORMALIZATION FOR THE APPLICATION OF NEURAL NETWORKS TO COMPLEX INDUSTRIAL PROBLEMS

J. Sola and J. Sevilla
Department of Electrical and Electronic Engineering
Universidad Pública de Navarra. 31006 Pamplona, Spain

ABSTRACT

Recent advances in artificial intelligence have allowed the application of such technologies in real industrial problems. We have studied the application of backpropagation neural networks to several problems of estimation and identification in nuclear power plants. These problems have been often reported to be very time-consuming in the training phase. Among the different approaches suggested to ease backpropagation training process, input data pretreatment has been pointed out, although no specific procedure has been proposed.

We have found that input data normalization with certain criteria, prior to training process, is crucial to obtain good results, as well as to fasten significantly the calculations. This paper shows how data normalization affects the performance error of parameter estimators trained to predict the value of several variables of a PWR nuclear power plant. The criteria needed to accomplish such data normalization are also described.

INTRODUCTION

Recent advances in the field of Artificial Intelligence have allowed the application of a series of new technologies to actual industrial problems [1]. In particular, neural networks, in their different topologies, have been applied to solve problems from very different fields [2].

In nuclear power plants the application of neural networks technology to different problems of identification [3,4] and estimation [5,6] has been reported. In the use of backpropagation networks to such kind of problems, several authors report very long training times and propose different approaches to solve this problem [7,8,9,10].

When using backpropagation networks, depending on the activation function of the neurons, it will be necessary to perform some pretreatment of data used for training. Supposing that logistic sigmoids are used, the interval of variation of the output variables has to be accommodated to the maximum output range of the sigmiod, that is from zero to one. Although such kind of normalization has been used [11], no specific procedure is clearly established in the literature.

An adequate normalization, not only for the network output variables but also for the input ones, previous to the training process is very important to obtain good results and to reduce significantly calculation time. In the work presented in this paper we have studied systematically the effect of different data normalization procedures in the performance of estimation networks trained with this data. The networks used have been designed to estimate the value of different variables of a PWR type nuclear power plant.

SYSTEM STUDIED

The results presented here are part of a wider project aiming to obtain virtual measurements based on neural networks for PWR nuclear power plants [12]. Particularly two cases have been studied, first an estimator of pressurizer pressure, and second an estimator of the power transferred between the reactor coolant system and the main vapor system.

In both cases supervised training with multilayer perceptron are used [2]. Data needed to train the networks as well as to evaluate their performance were obtained from simulations accomplished with the Nuclear Plant Analyzer of Westinghouse based on the code TREAT. Virtual measurement system was intended to work properly (in a first stage) around nominal power, thus several transients (ten) were selected, all starting in a normal state at 100% power, and simulated. These transients are summarized in Table I. The first 5 minutes of each transient, recording data each second, provided more than 3000 points, representing plant states, that were used for the networks development.

Load rejections (5%, 10% and 15%).
Reactor trip.
Safety injection actuation.
One train of main feedwater preheaters loss.
LOCA (0.5", 1" and 2" of equivalent diameter)
One pressurizer relief valve blocked open.

Table I. Transients simulated to obtain data.

The network studies have been carried out with the neural network simulator SINAPSIS. This PC-based simulator has been developed in the Universidad Pública de Navarra [13]. In both cases three-layer perceptron have been used, and trained through adaptive backpropagation algorithm [14]. Neuron activation function is, in all cases, a unity slope logistic sigmoid.

Input variables were a small number of typical plant variables (or combinations of them). In the case of the pressurizer estimator, 6 input variables have been used (see Table II) while in the power transfer case 8 variables were needed. Network architecture influence has been studied by evaluating the performance of a wide range of possibilities. Results showed that the optimum architecture varies with data pretreatment procedure. For the best normalization, relatively small networks perform as well as more complex ones. In order to avoid confusion in the discussions, all the results presented in this paper have been obtained with the same network architecture, namely 6-5-1 and 8-7-1, one for each of the cases studied. In both cases the training set was made with 100 points randomly selected out of the around 3000 available.

Mean temperature (loop 1). Pressurizer level. "Water in" (CVCS inlet flow + safety injection flow) "Water out" (CVCS letdown flow + pressurize safety valves flow + pressurizer relief valves flow + flow through brakes). Nuclear power (NIS) Turbine demand (pressure in the turbine first stage impulse chamber)

Table II. Input variables used for the pressure estimator.

RESULTS

When facing original data for the first time one realizes that they are expressed in very different units, not even belonging to the same units system. Not only the units are different, but also the order of magnitude of their absolute values are very different, ranging from 2.6 10^6 kw for nuclear power to turbine demand that is expressed normalized to unity. In Table III, a summary of the units and nominal values (at 100% power) for the variables of table II together with their maximum variation range along the transients considered are presented. Numerical values of the order of 10^6 generate extremely large numbers when calculating network output, exceeding ordinary computers capacity. So, it is completely necessary to perform some kind of input data pretreatment.

VARIABLE	UNIT	NOMINAL VALUE	MAXIMUM VARIATION RANGE
MEAN TEMPERATURE	ºC	308.95	312---- ----290
PRESURIZER LEVEL	%	57.89	64---- ----0
WATER IN	Kg/sg	3.78	41---- ----0
WATER OUT	Kg/sg	4.53	250---- ----0
NUCLEAR POWER (NIS)	kw	$2.686*10^6$	$3*10^6$---- ----0
TURBINE DEMAND	$^0/_1$	1	1---- ----0
PRESURIZER PREASURE	Kg/cm^2	157.1	175---- ----62

Table III. Summary of the units and nominal values at 100% power for the variables of table II.

To investigate the effect of input data pretreatment we have first selected one training set (of 100 points randomly selected). With this original set we have created 5 different sets with different normalizations of the same data. This "normalizations" are linear transformations of the original data. The transformations used are shown in table IV. All data transformations, different than the one labeled 5, are not normalizations strictly speaking, and are artificially introduced in order to study systematically the effect of input data variation range in the output performance of the network.

The transformations labeled #1, #2 and #4 (see table IV) are accomplished though product of input data with specific factors for each variable. In all cases this factors are well below unity in order to obtain absolute numbers well suited for computation. In the first “normalization” (second column of table IV), factors are 10^{-4} and 10^{-5}, so that the big differences in the absolute values of the original data are almost kept. Comparing this case with the others can be studied if this differences in the absolute values are as interesting for network performance as they are for human interpretation (the origin of this values are the unit systems selected by plant design engineers). Normalizations #2 and #3 are introduced to evaluate the effect of “equalizing” absolute values of the variables. With transformation #2, the original difference of 10^6 between the highest and the lowest data is reduced to 10^4, and with #4 to 10^3.

Transformation #3 is performed dividing each data by the nominal value of the corresponding variable at 100% of nuclear power. Absolute variable variation obtained is of three orders of magnitude, similarly as in #4 but through a different procedure. This allows to study whether variations in the network performance are due only to data absolute values variability.

Finally, with normalization #5 it is intended to expand input data interval of variation to the input layer neurons interval of variation. As these are unity slope sigmoids, this interval runs from 0 to 1. Nevertheless, the maximum output range used in normalization #5 (see table IV) runs from 0.05 to 0.95. This output range has been selected after investigating the performance of three different intervals: (0.1, 0.9), (0.05, 0.95) and (0, 1). The final average mean errors obtained are very similar in all cases (differences below 0.05%), while the one chosen shows slightly better performance in the initial stages of the training phase (up to $5. 10^4$ iterations).

In all cases the output variable is scaled following criteria for normalization number 5. That is: $((P_{RCS}/157,15)-0.4)$ 1,285 + 0.05. Remember, from table III, that P_{RCS} is expressed in Kg/cm^2 , and 157,1 is the value at 100% power. In the transients under consideration this variable never falls below 40%, this is included by subtracting 0,4. The factor 1,285 is this scaling factor that assures that variations in P_{RCS} will become in an almost unity variation. This procedure assures that all values can be provided by the output neuron (unity slope sigmoid).

NORMALIZATION	1	2	3	4	5
	FACTOR	FACTOR		FACTOR	TRANSFORMATION
MEAN TEMPERATURE	10^{-5}	10^{-4}	M.T.*	10^{-4}	((M.T.*)-0.94)*12.587)+0.05
PRESURIZER LEVEL	10^{-5}	10^{-4}	P.L.*	10^{-3}	((P.L.*)*0.8181)+0.05
WATER IN	10^{-4}	10^{-4}	A. IN*	10^{-2}	((A.IN*)*0.08181)+0.05
WATER OUT	10^{-4}	10^{-4}	A OUT*	10^{-2}	((A.OUT*)*0.01636)+0.05
NUCLEAR POWER (NIS)	10^{-5}	10^{-6}	N.P.*	10^{-5}	((N.P.*)*0.75)+0.05
TUBBINE DEMAND	10^{-4}	10^{-4}	T.D. *	10^{-2}	((T.D.*)*0.9)+0.05
Orders of magnitude of difference	6	4	3	3	0

(*) Value of each variable divided for the value of the variable to the 100% of power.

Table IV. Different data transformations used as normalizations.

With this data, the same network (6-5-1) was trained in an identical manner. Its performance was quantified averaging the differences between network output and real output for all available data, that is more than 3000 points (including the 100 used for training). This mean error (expressed in % of nominal value, i.e. 157.15 Kg/cm^2) is plotted in figure 1 for the normalizations considered. In the five cases the mean error after 10, 20, 100 and 200 thousand training iterations it is represented.

In figure 2 the results obtained from a similar study performed with the network designed to estimate power transferred between RCS and main vapor system, are presented.

In both cases it can be seen that error gradually decreases as normalization reduces the differences between the variation range of the different variables. It is interesting to note that when the variables span in orders of magnitude is the same (as happens with normalizations 3 and 4) the results are better for the case with less variables out of the most common order of magnitude. From the worst to the best case, results are better in one order of magnitude.

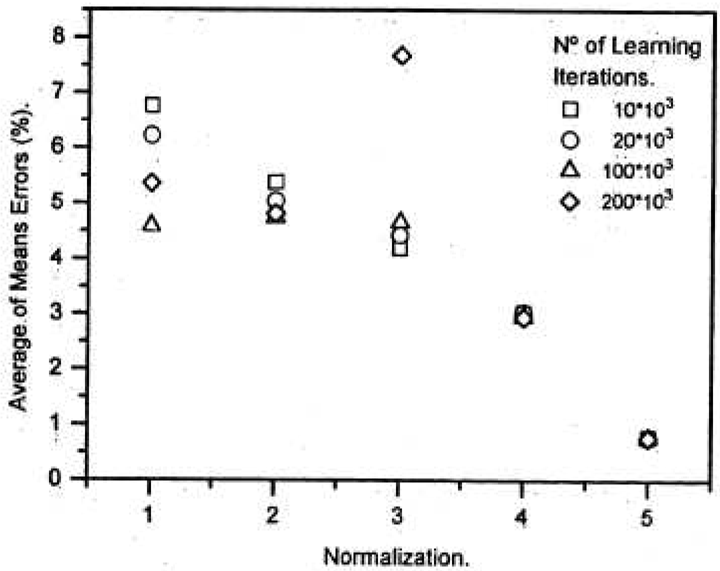

Fig. 1. Average mean error of the pressure estimation network for different normalizations.

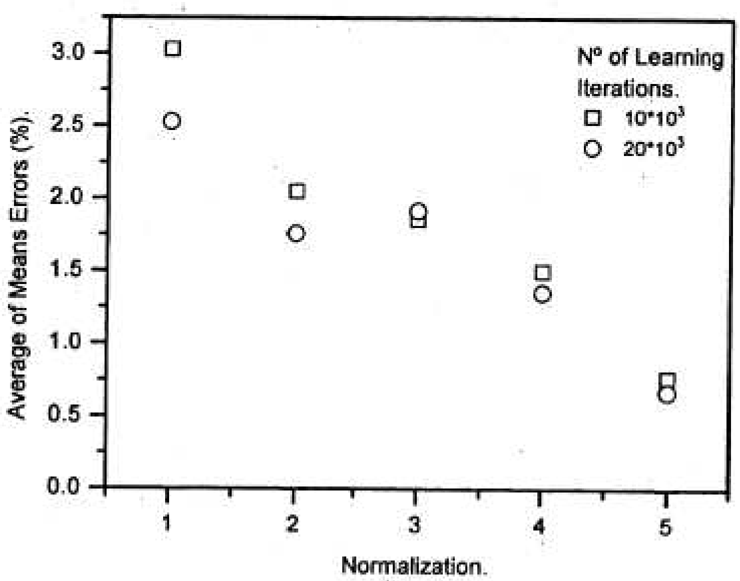

Fig. 2. Average mean error of the transferred power estimation network for different normalizations.

Most striking is the fact that errors do not follow a constant decrease with the number of iterations, but on the contrary they reach a certain minimum and start increasing again as iterations grow. This can be appreciated more clearly in figure 3, where the same mean error data are plotted against the number of iterations with different curves for each normalization. The worsening of the results as training iterations progress is more evident for normalizations 1 and 3.

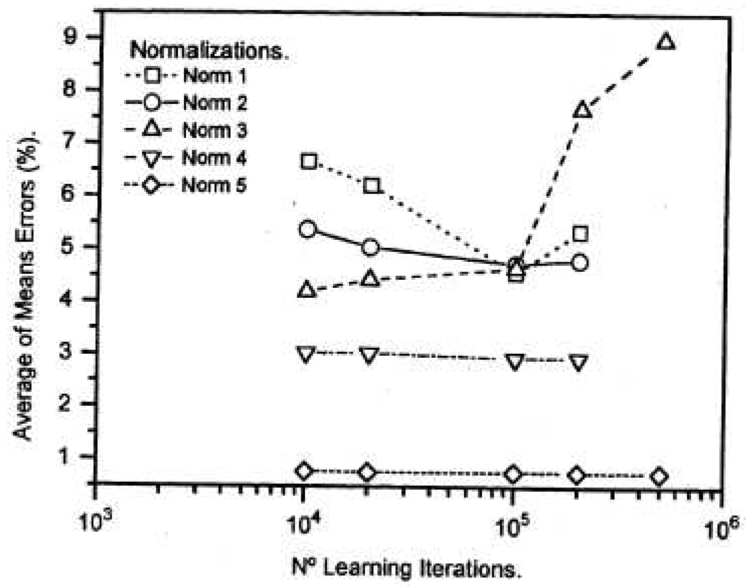

Fig. 3. Average mean error evolution for different numbers of training iterations (pressure estimation network).

DISCUSSION

The results previously shown can be summarized in two relevant features. On one hand, network performance (training iteration number and final error attained) is enhanced as input variable ranges are “equalized” by normalization. On the other hand, when the iteration number exceeds a certain value, the square mean error tends to rise again; and this value depends on the normalization method.

A possible explanation for the second observation can be found in the well known effect of "overtraining" [15], what means that the network tends to learn "by heart" the training set. Thus, as training data are fitted more closely, its performance for other data start to become worse. This overtraining process can be expected when the training set size is smaller than (approximately) ten times the number of connections of the network as is the case in the problems under study. Results obtained training the networks with training sets of different number of elements strongly support this idea.

To understand why input data normalization can enhance network performance, the first point is to remember that the neural network simulator used, initializes weights to random values in the (-1, 1) interval. The slope of the sigmids used as activation functions is also unity. On the other hand, all normalizations considered are linear scale transformations and thus, the minimum to which the network converges should be the same one in all cases, only shifted by the same linear transformation. Therefore, the initial state for backpropagation algorithm to begin is always a point in the vicinity of coordinate space origin, while distance to the desired minimum is drastically changed by the scales considered in each case. So, scales that compress all the searching space to a unitary hypercube reduce the distance to be covered, iteration by iteration, by the backpropagation algorithm.

Furthermore, if scales are very dissimilar for the different values, the bigger ones will have a higher contribution to the output error, and so, the error reduction algorithm will be focused on the variables of higher values, neglecting the information from the small valued variables. We have tried to test this idea by tracking weights shift as backpropagation proceeds in different cases. Results obtained support this explanation.

Previous discussion indicates that the best situation should be when all input variables are in the same order of magnitude, and specially if they are in the order of one, as happens with the best normalization presented in the study. However, to perform this transformation one has to define the absolute maximum and minimum values for each input variable. An exact extreme value assignment can be difficult in many real problems. Nevertheless, to obtain a good normalization no special accuracy is needed. Extreme value variations of 20 or 30 % do not alter significantly the results obtained. We think that in most cases there will always be an heuristic argument to assign extreme values within a $\mp$ 30% interval. For example, when considering pressure of a PWR plant, nominal value is 157.15 Kg/cm^2 ,and design maximum pressure is around 180 Kg/cm^2. So, any value between 160 and 200 can be taken for the “maximum” to perform the normalization with approximately the same results.

CONCLUSIONS

A systematic study on the effect of input data pretreatment prior to its use for neural network training has been accomplished. The study has been performed with two neural networks developed to estimate two PWR nuclear power plant variables.

We can conclude that an adequate normalization of input data prior to network training generates two clear advantages:

- Estimation errors can be reduced in a factor between 5 and 10.
- Calculation time needed in the training process to obtain such results is reduced in one order of magnitude.

An adequate normalization is a linear scale conversion that assign the same absolute values to the same relative variations.

ACKNOWLEDGMENTS

The authors acknowledge Westinghouse Spain, and specially Mr. Carlos Pulido, for their assistance with TREAT-NPA data generation. Fruitful discussion with Dr. G. Lera is also recognized.

REFERENCES

[1] S. H. Rubin Ed. “Artificial Intelligence for Engineering, Design, and Manofacturing”, ISA Transactions Vol. 31, 2 (USA, 1992)

[2] R. P. Lippmann “An Introduction To Computing With Neural Nets”, IEEE ASSP, 4, 2, 4. (USA, 1987)

[3] R. Kozma, M. Kitamura, J.E. Hoogemboom, M. Sakamura. “Credibility of Anomaly Detection in Nuclear Reactors Using Neural Networks”, Transactions of the American Nuclear Society, Vol 70, p 102-3 (USA, 1994)

[4] B.C. Hwang. “Intelligent control for Nuclear Power Plant Using Artificial Neural Networks”, Proc. IEEE International Conferencd on Neural Networks, Vol 4 p. 2580-4 (USA, 1994)

[5] K. Nabeshima, R. Kozma, E. Turkcan. “Analysis of the Respose Time of An On-Line Boiling Monitoring System in Research Reactors With MTR-type Fuel”, Proc. Of 16th International Meeting on Reduced Enrichment for Research and Test Reactors, p. 266-73 (Japan, 1993)

[6] O. Sung-Hun, K. Dae-Il, K. Kern-Jung. “A Syudy on The Signal Validation of The Pressurizer level Sensor in Nuclear Power Plants”, Transactions of The Korean Institute of Electrical Engineers Vol 44, 4, p. 502-7 (S. Korea, 1995)

[7] W.J. Kim, S.H. Chang, B. H. Lee. “Applications of Neural Networks to Signal Prediction in Nuclear Power Plant”, IEEE Transactions on Nuclear Science Vol 40, 5, p. 1337-41 (USA, 1993)

[8] J. E. Vitela, J. Reifman. “Improving Learning of Neural Networks for Nuclear Power Plant Transient Classifications”, Transactions of The American Nuclear Society Vol 66, p. 116-17 (USA, 1992)

[9] J. Reifman, J. E. Vitela. “Accelerating Learning of Neural Networks With Conjugate Gradients for Nuclear Power Plants Applications”, Nuclear Technology Vol 106, 2, p. 225-41 (USA 1994)

[10] K. K. Shukla. “A Neuro Computer Based Learning Controller for Critical Industrial Applications”, Proc. Of 1995 IEEE/IAS Int. Conf. On Industrial Automation & Control, p. 43-8 (USA, 1995)

[11] B. R. Upadhyaya, E. Eryurek. “Application of Neural Nets for Sensor Validation and Plant Monitoring”, Nuclear Technology Vol 97, p.170-6 (USA, 1992)

[12] J. Sevilla, J. Sola, C. Achaerandio, C. Pulido. “Neural Network Technology Applications to Nuclear Power Plants Instrumentation”, Proc. Of 22 Sociedad Nuclear Española Anual Meeting (in press)

[13] G. Lera, personal communication. Dr. Lera can be contacted at glera@si.upna.es.

[14] A.G. Parlos, J. Muthusami, A. F. Atiya. “Incipient Fault Detection And Identification In Process System Using Accelerated Neural Network Learning”, Nuclear Technology Vol 105, 2 (USA, 1994)

[15] J. Sjöberg, L. Ljung, “Overtraining, Regulation, and Searching for Minimum in Neural Networks”. Linköping University, Departament of Electrical Engineering Internal Repport. sjoberg@isi.liu.se