THE EXTRACTION OF KNOWLEDGE RULES FROM ARTIFICIAL NEURAL NETWORKS APPLIED IN THE ELECTRIC LOAD DEMAND FORECAST PROBLEM

How Artificial Neural Networks Retain Knowledge and Make Reliable Forecasts

Tarcisio R. Steinmetz, Adelmo L. Cechin and Jose V. Canto dos Santos
PIPCA - UNISINOS, Av. Unisinos, Sao Leopoldo, Brazil

Keywords: Rule extraction from Artificial Neural Networks, Fuzzy Set Theory, Principal Components Analysis, Electric Load Demand Forecast.

Abstract: We present a methodology for the extraction of rules from Artificial Neural Networks (ANN) trained to forecast the electric load demand. The rules have the ability to express the knowledge regarding the behavior of load demand acquired by the network during training process. The rules are presented to the user in an easy to read format, such as *IF premise THEN consequence*. Where *premise* relates to the input data submitted to the network (mapped as fuzzy sets), and *consequence* appears as a linear equation describing the output to be presented by the network, should the premise part holds true. Experimentation demonstrates the method's capacity for acquiring and presenting high quality rules from neural networks trained to forecast electric load demand for several amounts of time in the future.

1 INTRODUCTION

One important issue concerning the requirements of proper load demand forecast methods is the ever increasing dependency of electricity supply for today's industrial societies. Hence, the last decades have shown large investments from energy supply companies in order to improve operation security of electric networks and to ensure quality of service of energy supply for the costumers (Ghods and Kalantar, 2008). These objectives could be achieved through the use of a better knowledge of the load demand behavior for the area supplied by energy supply companies. Such knowledge can even be used to guide the company's tactical and strategic decision making within the company's administrative areas.

This work presents a methodology designed for the extraction of rules form Artificial Neural Networks trained to forecast electric load demand for several amounts of time in the future. The rules obtained describe the knowledge acquired by the network during the training phase. The rules provide insight about the load demand behavior for the area where the training data have been gathered (a city, for instance), such as the impact that each of the input variables cause on the load demand, under what circumstances occurs drastic changes in the load demand pattern, among other important information to support tactical and strategic decisions throughout the energy supply company. This paper proceeds as follows: in the next section we discuss some theoretical aspects. Section 3 details FAGNIS, the rule extraction method used in this work. Section 4 demonstrates the methodology proposed for the proper rule extraction from the trained ANNs. Section 5 shows some of the experiments used to validate the method and the results obtained. In Section 6 we finish the document, presenting our conclusions.

2 THEORY

This section deals with the theoretical concepts used in this paper. Fuzzy Set Theory and Principal Components Analysis are used in this work, however, due to space limitations they are not covered here. The reader should refer to (Angelov, 2002) and (Hastie et al., 2009) to read about these topics.

2.1 Electric Load Demand Forecast

The electricity demand or system load encompasses the summation of electric usage at each consumption point (users) supplied by an electric supply facility.

R. Steinmetz T., L. Cechin A. and V. Canto dos Santos J. (2009).
THE EXTRACTION OF KNOWLEDGE RULES FROM ARTIFICIAL NEURAL NETWORKS APPLIED IN THE ELECTRIC LOAD DEMAND FORECAST PROBLEM - How Artificial Neural Networks Retain Knowledge and Make Reliable Forecasts .
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Intelligent Control Systems and Optimization*, pages 195-200
DOI: 10.5220/0002198201950200

Its behavior is highly dynamic and difficult to comprehend. The amount of variables involved in the characterization of the load demand curve is large indeed, and different effects are perceived by the same variables in different regions of the globe. However, works such as (Gross and Galiana, 1987), (Srinivasan et al., 1995), (Srinivasan et al., 1999) and (Ghods and Kalantar, 2008) show that certain factors are commonly responsible for affecting the load demand.

2.2 Rule Extraction from Trained Neural Networks

The reason for the successful application of AANs in fields as diverse as academia, industry and commerce is its generalization capabilities. However, this high power of generalization comes at a price: it prevents the network from expressing the knowledge acquired during the training phase. Thus the network can be seen as a black box, presenting to the user the predicted output based on the input data, while important knowledge about the problem studied remains encrypted within the network's weight matrix, never to be discovered.

In order to solve this problem, rule extraction techniques can be used to acquire the knowledge embedded in the network's weight matrix, and then to present it to the user in a clear interface. As mentioned before, rules have a IF *premise* THEN *consequence* structure, where *premise* somehow defines the vector of input data presented to the network and *consequence* describes the output to be obtained should the *premise* part holds true. Mostly of the rule extraction methods used today rely on Fuzzy Sets concepts to describe the premise part of the rules (Cechin, 1998), (Benitez et al., 1997).

In an important survey concerning several rule extraction methods, Andrews et al. (Andrews et al., 1995) present several interesting features displayed by transparent neural networks, that is, ANN capable of describing their knowledge to the user. In this work the authors mention that ANNs with explanatory capabilities are capable of (among others): (1) operating in conjunction with symbolic intelligent systems; (2) controlling critical applications such as air traffic control and support scientific theory formulation.

3 FAGNIS - RULE EXTRACTION FROM SIGMOID NETWORKS

This section describes FAGNIS (Cechin, 1998), the rule extraction method selected for this work. FAGNIS has been considered because of its ability to extract rules from standard feedforward neural networks, with or without shortcut connections, and heaving one or more hidden layers. Further, since FAGNIS performs on an already trained network, it has no dependency on its training algorithm. In fact, any training algorithm can be used, from the standard backpropagation to algorithms yet to be created. Requirements such as special ANN architectures and special or adapted learning algorithms are mandatory in the majority of rule extraction methods, namely (Jang et al., 1997) and (Nauck et al., 1994).

FAGNIS begins its extraction procedure by splitting the sigmoid curve within the hidden neurons in three regions. These regions are then transformed in straight lines, which are mapped by very simple equations, as illustrated in Figure 1.

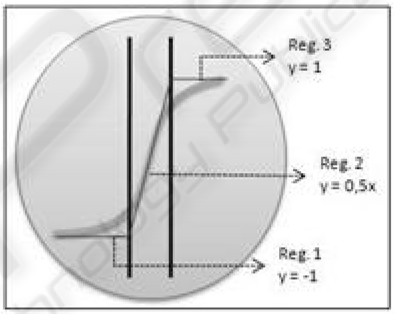

Figure 1: Separation of the sigmoid curve within the hidden neurons performed by FAGNIS.

Next, the training data are once more submitted to the network, where FAGNIS verifies the resulting activation of the hidden neurons for each of the data points. The data points are then grouped according to the activation regions (as shown in Figure 1) generated within the network's hidden neurons.

To assemble the premise part of the rules, FAGNIS transforms each group of data points found in the previous step in fuzzy sets. The fuzzy sets are represented by the midpoint of each group. The consequence part of the rule is defined as a linear equation that represents the output dependence of the network on the input data. The expressions below show two rules acquired from a fictitious neural network

IF (x_1, x_2) is G_1 THEN $y = x_1 w_{ij} + x_2 w_{ij} + k$

IF (x_1, x_2) is G_2 THEN $y = x_1 w_{ij} + x_2 w_{ij} - k$

where G_1 and G_2 are fuzzy sets (with membership functions μ_1 and μ_2 respectively), w_{ij} is the weight linking the i-th neuron to the j-th neuron and k is the intercept value for the equation.

4 METHOD

The first step of the process is to prepare the ANN to be used as the forecaster model. Some consolidated techniques, such as variable selection method and cross-validation technique were used to improve the model, as well as to decide on key issues concerning the ANN's architecture. The Mean Absolute Percentage Error (MAPE) metric was selected to measure the ANN's accuracy, as shown in Equation (1):

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \qquad (1)$$

where n is the number of data points, A_i is the actual value and F_i is the forecast value. Once the dataset is composed, each data point is normalized as shown in Equation (2):

$$nd_i = \frac{d_i - \mu_D}{\sigma} \qquad (2)$$

where nd_i is the normalized data, d_i is the actual data (from the dataset), μ_D is the mean of the data column and σ is the standard deviation for the data column.

Principal Components Analysis (PCA) is then executed on the dataset, and the resulting dataset is then used to train the neural network. The principal components are selected based on Jolliffe's criterion (Jolliffe, 2002).

Once the neural network is properly built and trained, FAGNIS can be used to extract knowledge rules from it.

The algorithm is executed as detailed in Section 3. Once the execution is terminated, the rules can be analyzed and interpreted. The equations on the consequence part of the rules explain the load demand behavior for the data assigned to the fuzzy sets described in the premise part. To determine which of the input variables is the most important, that is, the one that has the major influence on the load demand, the user needs simply to identify which of the independent variables of the equation has the highest absolute coefficient value.

5 EXPERIMENTS AND RESULTS

This section describes the results obtained from two experiments used to evaluate the proposed methodology. The first experiment concerns the extraction of rules from ANN trained to predict the average load for the next hour. In the second experiment, the forecast window is expanded to the next month.

5.1 Experiment 1

The load demand forecast for the next hour constitutes a classical problem within this field, and the adoption of neural networks techniques usually leads to excellent results. Regardless of the triviality of this problem, the energy supply companies should not underestimate the value of such information: the load demand for the next hour provides support to several of the company's tactical decisions like the expansion of transmission lines, equipment maintenance schedule and other routine activities.

The dataset used in this experiment corresponds to the hourly load recorded from 2003 to 2007, For a capital city holding approximately 1.4 million of citizens. The data has been arranged in such a way that daily, weekly, monthly and annually load patterns could be learned by the neural network. The list below represents the structure of the dataset prior to PCA application:

- load demand for the last twenty-four hours (24 columns),
- load demand for the forecast hour registered in the last six days (6 columns),
- load demand for the same day and forecast hour registered in the last three weeks (3 columns),
- load demand for the forecast hour and day registered in the last month (1 column),
- dependent variable: average load demand for the next hour.

After data processing, the principal components were extracted. Jolliffe's Criterion informed that the first eight components should be used as the input layer of the neural network. The data belonging to the remaining components were discarded from the experiment.

The neural network selected by tenfold cross-validation method has eight neurons in the input layer, four in the hidden layer and one in the output layer. The network's accuracy, measured by the MAPE metric, is of 0,027%.

Some of the rules extracted by FAGNIS appear in table 1. Not all the thirty-four rules were displayed due to space reasons. The rules show that the three first principal components have an increasing effect on the load (their coefficient values are positive in all the thirty-four rules). On the other hand, the fourth principal component has a decreasing effect on the load demand (it has negative coefficient value in all the rules).

The rules are arranged in the following format:

IF (PC_1 is F_1) AND (PC_2 is F_2) ... AND (PC_8 is F_8)

THEN $y = int + KPC_1 + KPC_2 + \ldots + KPC_8$

Table 1: Some of the rules found by FAGNIS in experiment 1.

Rule #	Rule description	Data points
1	IF x = (0.319 -1.272 -0.431 -0.091 0.275 -0.349 -0.159 -0.312) THEN y = (0.036 0.189 0.174 0.128 -0.175 0.159 0.047 0.038 -0.093)	8946
2	IF x = (1.913 1.350 0.327 -1.651 -0.609 0.923 0.301 0.317) THEN y = (-0.004 0.208 0.191 0.179 -0.096 0.204 -0.001 0.237 0.131)	3948
3	IF x = (1.179 -0.652 1.299 0.420 0.080 0.006 0.867 0.390) THEN y = (-0.085 0.215 0.207 0.156 -0.183 0.136 0.111 0.192 0.162)	3556

where PC_n is the principal components used as the input layer of the neural network, F_n are the fuzzy sets representing the data being submitted to the ANN's input layer, K are the coefficient values of the linear equation and *int* is the point where the straight line defined by the equation intercepts the Y axis.

Figure 2 depicts the first column of the rotation matrix resulted from PCA application. It says that the first principal component is composed mainly by the load of the forecast hour registered one day ago. Table 2 presents the results of the same analysis for the remaining principal components.

Table 2: Most important variables used for principal components characterization.

PC	Description
1	Load for the forecast hour, 1 day ago
2	Load for the forecast hour, 5 days ago
3	Load registered 24 hours ago
4	Load for the forecast hour, 1 week ago
5	Load registered 14 hours ago
6	Load registered 11 hours ago
7	Load for the forecast hour, 4 days ago
8	Load for the forecast day and hour, 1 month ago

Based on the rules found and the information detailed on table 2, the following assertions can be made:

1. the load demand registered in the last twenty-four hours before the forecast, as well as one day before and five days before have an increasing effect on the load demand
2. the load demand for the same time of the forecast, registered one week ago, decreases the load demand for the next hour

5.2 Experiment 2

The load demand forecast for the next month is a task much more difficult than that of the previous experiment. As the window of forecast expands to such a long time, economic factors begin to play a more important role in shaping the load demand curve (Srinivasan et al., 1999). The load demand for the next month consists in strategic information to energy supply companies. It supports the company to purchase an amount of energy very close to the amount to be used by its costumers, thus increasing the company's profit.

In this experiment, electric demand and climatic data were used to determine the load demand for the next month for a small city with a large number of industries. The data have been stored on a daily basis, for the period of 2005 to 2007. The file structure before PCA is shown below:

- residential load demand registered for 120, 90, 60 and 30 days prior to forecast (4 columns),
- industrial load demand registered for 120, 90, 60 and 30 days prior to forecast (4 columns),
- commercial load demand registered for 120, 90, 60 and 30 days prior to forecast (4 columns)
- average temperature registered for 120, 90, 60 and 30 days prior to forecast (4 columns),
- average relative air humidity registered for 120, 90, 60 and 30 days prior to forecast (4 columns),
- dependent variable: average load demand for the next 30 days.

The first seven principal components were used to build the training dataset to the neural network. The reason for the separation of the load in three different categories relies on the city's economic structure: it is mainly industrial. However, the commerce sector has shown rapid increase in the last decades. Thus, it is expected that the rules found show high dependency of industrial load to forecast the average monthly load for this city.

The neural network selected via tenfold cross-validation has seven neurons in the input layer, thirty-two in the hidden layer and one in the output layer. Shortcut connections were not used. This architecture resulted in a MAPE of 3.26%, however, due to the elevated quantity of hidden neurons, more than 300 rules were extracted. It means that many fuzzy

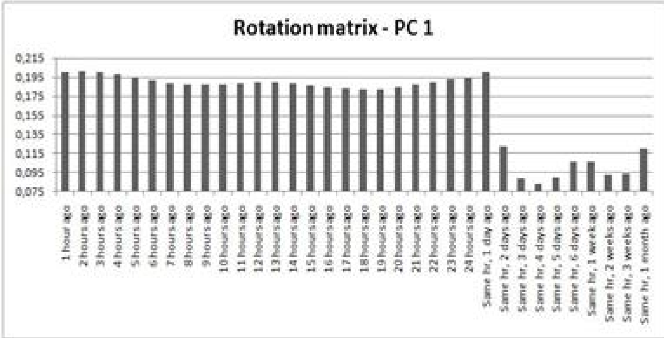

Figure 2: Quantity of information from original dataset used to create the first principal component.

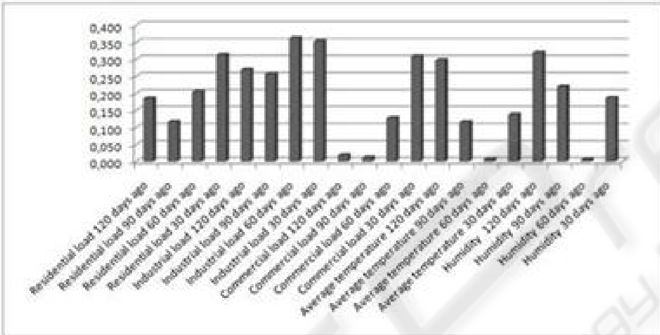

Figure 3: Quantity of information from original dataset used to create the first principal component.

sets were necessary to map the knowledge acquired by the network, and these sets refer to very few data points in the training data. To solve this problem, a new network structure was used in the rule extraction procedure: it has eight hidden neurons, and its MAPE value is of 6.53%. This neural network produced thirty rules.

Table 3 shows the most important rules extracted by FAGNIS. They can be read by the same manner as those shown in the previous experiment. Again, not all the rules could be presented due to space reasons. Figure 3 details the data on the first column of the rotation matrix resulted from PCA application. It shows that the industrial load registered sixty days ago has a strong relation to the load for the next month. However, it is clear that the industrial load registered thirty days ago has also a significant participation in shaping the overall monthly load. Table 4 presents the quantity of information given by the original dataset to build the remaining principal components.

Based on the analysis of the first rule found and the rotation matrix resulted from PCA application (Table 4), it is possible to verify that the industry load demand has high impact on the general monthly load demand of the considered city.

Figure 4 shows the monthly load demand curve for the city. The dots represent the data points associated with rule number 1. This shows that the Fuzzy Sets of the first rule have elevated energy consumption. The analysis described in this experiment can be replicated to the other rules, so that all knowledge learned by the neural network can be acquired.

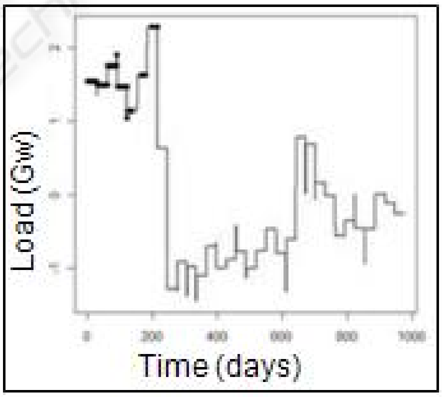

Figure 4: Monthly load demand and data explained by Rule 1.

Table 3: Rules found by FAGNIS in experiment 2.

Rule #	Rule description	Data points
1	IF x = (-2.863 -0.715 1.839 -0.294 0.264 -0.014 0.035) THEN y = (1.442 -0.024 0.001 0.032 -0.033 0.005 -0.004 -0.003)	110
2	IF x = (2.554 -1.616 -1.161 0.165 0.154 -0.095 -0.022) THEN y = (-0.290 -0.002 -0.378 -0.117 -0.392 0.006 -0.005 -0.235)	68
3	IF x = (4.356 0.069 2.064 -0.852 -0.152 -0.221 0.596) THEN y = (-2.804 0.874 0.339 -0.386 0.737 -0.027 0.001 0.127)	42

Table 4: Most important variables used for principal components characterization.

PC	Description
1	Industrial load, 60 days ago and Industrial load, 30 days ago
2	Average temperature, 60 days ago
3	Average temperature, 120 days ago
4	Industrial load, 30 days ago
5	Average humidity, 60 days ago
6	Average humidity, 90 days ago
7	Average humidity, 120 days ago

6 CONCLUSIONS

A methodology for the acquisition of rules from neural networks trained to forecast electric load demand has been presented here. Results found through several experiments (been two of them shown in this paper) attest the methodology's efficiency in extract and present high quality rules for different amounts of time in the future.

Throughout the execution of many experiments, it was made clear that there is a need to differentiate the neural networks of load forecast from those used to rule extraction: the former needs several training cycles in order to obtain a perfect fit to the load demand curve; the latter requires only a few training cycles to obtain the overall knowledge about the load demand, that is, so that a small number of rules can be used to refer to a large quantity of data points.

Both the forecast model and the rules acquired can be used as decision support tools for energy supply companies. For example, several simulations could be used for the executives to better understand load demand behavior in different scenarios, such as future climatic changes.

REFERENCES

Andrews, R., Diederich, J., and Tickle, A. (1995). Survey and critique of techniques for extracting rules from trained neural networks. *Elsevier Knowledge-Based Systems*.

Angelov, P. (2002). *Evolving Rule-based Models: A Tool for Design of Flexible Adaptive Systems (Studies in Fuzziness and Soft Computing)*. Physica-Verlag, Heidelberg, first edition.

Benitez, J., Castro, J., and Requena, I. (1997). Are artificial neural networks black boxes? *Neural Networks, IEEE Transactions on*.

Cechin, A. (1998). *The Extraction of Fuzzy Rules from Neural Networks*. Shaker Verlag, Tubingen.

Ghods, L. and Kalantar, M. (2008). Methods for long-term electric load demand forecasting; a comprehensive investigation. *Industrial Technology, 2008. ICIT 2008. IEEE International Conference on*.

Gross, G. and Galiana, F. (1987). Short-term load forecasting. *Proceedings of the IEEE*.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, New York, second edition.

Jang, J., Sun, T., and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing. A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, New Jersey.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics, New York.

Nauck, D., Klawonn, F., and Kruse, R. (1994). *Neuronale Netze und Fuzzy-Systeme*. Vieweg and Sohn.

Srinivasan, D., Chang, C., and Liew, A. (1995). Demand forecasting using fuzzy neural computation, with special emphasis on weekend and public holiday forecasting. *Power Systems, IEEE Transactions on*.

Srinivasan, D., Tan, S. S., Cheng, C., and Chan, E. K. (1999). Parallel neural network-fuzzy expert system strategy for short-term load forecasting: system implementation and performance evaluation. *Power Systems, IEEE Transactions on*.