One step ahead prediction using Fuzzy Boolean Neural Networks[1]

José A. B. Tomé
INESC-ID, IST
Rua Alves Redol, 9 1000 Lisboa
jose.tome@inesc-id.pt

João Paulo Carvalho
INESC-ID, IST
Rua Alves Redol, 9 1000 Lisboa
joao.carvalho@inesc-id.pt

Abstract

Time series prediction is a problem with a wide range of applications, including energy systems planning, currency forecasting, stock exchange operations or traffic prediction. Accordingly, a number of different prediction approaches have been proposed such as linear models, Feedforward Neural network models, Recurrent Neural networks or Fuzzy Neural Models. In this paper one presents a prediction model based on fuzzy rules that relate past data values with the next unknown value to be estimated. A Fuzzy Boolean Neural Network has been used for that purpose and the laser data set of the Santa Fe contest has been used for illustration purpose. The results turned to be encouraging when compared with other published methods.

Keywords: Time Series Prediction, Fuzzy Neural Nets.

1 Introduction

Many practical prediction problems, such as stock exchange, traffic or water stream flow forecast, are satisfied with a one step ahead prediction. That is, in this type of problems one aims to estimate a time series value v_t given past values v_{t-1} , v_{t-2},… , v_{t-k}, their slopes or other simple functions of these values. Traditionally, most techniques to model this type of time series assume linear relationships among the variables, as AR models based on the Box and Jenkins method [1], [5]. Non-linear models have also been used, mostly based on neural network architectures both feed forward [3] and recurrent [2], but also using fuzzy concepts [4].

In this paper one introduces a new prediction model based on Fuzzy Boolean Neural Networks (FBN), which have been previously been presented as Nets capable of learning qualitative rules and of reasoning using those rules [10], [11]. The idea behind this approach is that one-step ahead prediction is nothing more that reasoning about the next value from passed values, assuming the variable behaviour can be described by a set of qualitative rules.

In order to test such a conjecture one has chosen a well-known and studied problem, which is the estimation of a time series obtained from laser data, the so called data set A from the Santa Fe contest on prevision [14].

Section II presents the FBN’s architecture, with details about its learning and fuzzy reasoning capabilities. Section III addresses the application details of FBN’s to time series prediction and a study on the effect of various net parameters on the net performance for laser data example. Section IV concludes and discusses some issues that deserve further attention, such as the possibility to understand the Net results, through the fuzzy rules results.

[1] This work is partially supported by the FCT - Portuguese Foundation for Science and Technology under project POSI/SRI/47188/2002

2 Fuzzy Boolean Nets

Fuzzy Boolean Nets [10] are inspired on natural neural systems, in which neurons are of the type on/off and where activity of concepts/variables is given by the rate of activated (firing) neurons on the corresponding neural areas. In such nets every neuron of a consequent variable or concept has the capability to memorize an individual and very simple (binary) vision of each one of the possible qualitative rules. In this model fuzziness is an inherent emerging property that can be derived from the network behaviour. A very interesting property of the model concerns its robustness, in the sense that it is immune to individual neuron or connection errors, which is not the case of other models, such as the classic artificial neural nets.

The architecture of the model is anchored in cards/areas of neurons. Meshes of weightless connections between antecedent neuron outputs and consequent neuron inputs link antecedent to consequent areas. Individual connections are random and neurons are binary, both in which concerns inputs and outputs. Neuron's internal unitary memories are provided, however, with a third state in addition to the classical "1" and "0", with the objective of identifying "not taught" situations. To each area a concept or variable is associated and the "value" of that concept or variable, when stimulated, is given by the activation ratio of that area (that is, the relation between activated -output "1"- neurons and the total number of neurons). Basically, each consequent neuron samples the antecedent spaces using for each antecedent a limited number (say m) of their neuron's outputs as its own inputs, being this number, m, much smaller then the number of neurons per area. This means that for rules with N-antecedent/1-Consequent each neuron has N.m inputs. The operations carried out by each neuron are just the combinatorial count of the number of activated inputs from every antecedent. They are performed with the classic Boolean operations (AND, OR), using as operands the inputs coming from antecedents and the Boolean internal state variables that are established during a teaching phase. It has been proved [10] that, from these micro operations (that is, on individual neurons using local and limited information), it emerges a global or macro qualitative reasoning capability on the concepts, which can be expressed on the form of rules of type:

IF Antecedent1 is A1 AND Antecedent2 is A2 AND ….THEN Consequent is Ci,

In the expression, Antecedent1, Antecedent2,.., Consequent are variables or concepts and A1, A2; …, Ci are linguistic terms (or fuzzy sets) of these variables (such as, "small", "high", etc.).

The proof [10] is achieved through the interpretation made to the relationship between antecedent and consequent activation ratios. The consequent activation ratio is given by:

$$\sum_{k1=0}^{L-1}\ldots\sum_{kN=0}^{L-1}\left(\prod_{i=1}^{N}\sum_{qi\in S_{ki}}\binom{m}{qi}.pi^{qi}.(1-pi)^{m-qi}.pr(q1,...,qN)\right) \quad (1)$$

In this expression L is the number of subsets of counts (each one a set of consecutive integers $\subset \{0,1,\ldots, m\}$), Ski are those sets, qi represents every possible number of activated inputs on antecedent i belonging to the set indexed by ki, pi are the antecedent activation ratios of area i and pr(q1,….,qn) are activation ratios of the internal binary memories (measured through every neuron), which have been established during the teaching phase. It is supposed that each neuron has one element of binary memory (with additional capability of memorizing if it has or not been taught) per set of counts on the antecedent inputs, (each set represented by the set of $q1 \in S_{k1}$ activated inputs on antecedent1,..., and $qn \in S_{kn}$ activated inputs on antecedent n).

Taking the inner sum above to the outside and considering each count individually rather than on sets Ski, the following expression is obtained:

$$\sum_{i1=0}^{m}\ldots\sum_{iN=0}^{m}\prod_{j=1}^{N}\binom{m}{ij}p_j^{\,ij}.(1-p_j)^{m-ij}.pr(i1,..,iN) \quad (2)$$

This equation represents at the macroscopic/network level an emergent behaviour, which can be viewed as a fuzzy qualitative reasoning that has been obtained from the microscopic neural operations defined above. In order to establish this fuzzy reasoning, the algebraic product and the bounded sum are interpreted respectively as the t-norm and t-conorm fuzzy operations. To this purpose the equations above should be interpreted as follows [10]:

Input variables, the activation ratios pj, are fuzzified through binomial membership functions of the form $\binom{m}{ij} p_j^{ij}.(1-p_j)^{m-ij}$. The evaluation of the expression for a given p_j represents the membership degree of p_j in that fuzzy set.

The product of the terms, the $\prod_{j=1}^{N}$, represents the fuzzy intersection of the antecedents (i=1,N), by definition of the above t-norm.

Considering the consequent fuzzy sets as singletons (amplitude "1") at the consequent UD values p(i1,..,iN), it follows that the equations represent the defuzzification by the Centre of Area method [10].

During the learning phase the network is submitted to a collection of experiments that will set or reset the individual neuron's binary memories and will establish the pr(i1,...,iN) probabilities in the above expression (1). For each experiment, a different input configuration (defined by the input areas specific samples) is presented to each and every one of the consequent neurons. This configuration addresses one and only one the internal binary memories of each individual neuron. Updating of each binary memory value depends on its selection (or not) and on the logic value of the consequent neuron. This may be considered a Hebbian type of learning [6] if pre and post-synaptic activities are given by the activation ratios: p_j for antecedent area j and p_{out} for the consequent area. For each neuron, the m+1 different counts are the meaningful parameters to take into account for pre synaptic activity of one antecedent. Thus, in a given experiment, the correlation between posterior synapse activity (p_{out}) and pre synaptic activity -the probability of activating a given decoder output $d(i_1,..i_N)$, corresponding to i_1 activated inputs on antecedent area 1, …, i_N activated inputs on area N, - can be represented by the probability of activating the different binary unitary memories. In practical terms, for each teaching experiment and for each consequent neuron, the state of binary memory $ff(i_1,...,i_N)$ is decided by, and only by, the Boolean values of decoder output $d(i_1, ..., i_N)$ and of the output neuron state considered.

Considering then the pr(k1,...,kN),in expression (2), as the synaptic strengths, one may have different learning types, depending on how they are updated. Here, one is considering the interesting case when non-selected binary memories maintain their state and selected binary memories take the value of consequent neuron, which corresponds to a kind of Grossberg based learning. It corresponds to the following updating equation (where indexes are not represented and p is used in place of pr, for simplicity) and where P^a is the probability of activating, in the experiment, the decoder output associated with p and p_{out} the probability of one consequent neuron to be activated:

$$p(t+1)-p(t) = P^a. (p_{out} - p(t)) . \qquad (3)$$

It can be proved [11] that the network converges to any taught rule. More specifically,

it may be concluded that with a set of coherent experiments -teaching the same rule- it will approach the taught value P_{out} proportionally to the distance between the present value of p and P_{out} itself, that is, with approaching zero decreasing steps.

Moreover, it has been proved [12] that the Net is capable of learning a set of different rules without cross-influence between different rules, and that the number of distinct rules that the system can effectively distinguish (in terms of different consequent terms) increases with the square root of the number m. This is an interesting result, since it should be verified in animal brains, that is, the number of synapses coming from a given concept area is expected, according this simple model, to be at least of the order of the square of the different linguistic terms on the consequent.

Finally, it has also been proved that this model is a Universal Approximator [11], since it theoretically implements a Parzen Window estimator [9]. This means that these networks are capable to implement (approximate) any possible multi-input single-output function of the type: $[0,1]^n \rightarrow [0,1]$.

These results give the theoretical background to establish the capability of these simple binary networks to perform reasoning and effective learning based on real experiments. The model can also be viewed as a natural application of the paradigm of computing with words. Since the nets present topologic similarities with natural systems and present also some of their important properties, it may be hypothesized that it may constitute a simple model for natural reasoning and learning.

Also, as the emergent reasoning may be interpreted as fuzzy reasoning, it may also be hypothesized that natural reasoning is intrinsically fuzzy.

3 Laser Data Prediction

The time series problem here treated is the well-known laser data of the Santa Fe contest [14]. This is data is approximately a Lorenz model of a two level system [7] and is considered an example of a stationary physical system, which behaviour and equations are well established.

During the learning phase values v_{t-1} , $v_{t-2},\ldots$, v_{t-k} of the variable on the data set are provided as antecedents and v_t is also provided as the consequent. The FBN is a black box forecast system that eventually acquires the governing qualitative rules of the time series. In the application phase, when values $v_{t'-1}$, $v_{t'-2},\ldots$, $v_{t'-k}$ are given as inputs it is supposed that $v_{t'}$ is predicted by the FBN.

In order to use the Boolean fuzzy net it is necessary to comply the continuous values of the antecedents with the binary inputs expected. Since the binary nature of antecedent inputs result from the fact that they come from an area of neurons where the respective variable value is the activation ratio, one has to define the interval limits for each variable and to find the ratio between the value of each continuous observation and the interval. Then a random assignation of "1"s and "0"s is made to every antecedent neuron in order to obtain the desired antecedent ratio. A Monte Carlo type method can be used for this purpose. That is, for each antecedent A_k one gets:

$$Oi^j = Rand(Vi^j/(Max-Min)) ;$$

where Oi^j is the output from neuron **i** in experiment **j**, Vi^j is the actual continuous value of the variable **k** in the same experiment and Rand(x) is a function that returns "1" with probability x or "0" otherwise.

The theoretical conditions for convergence [12] state that the number of inputs per consequent neuron and per antecedent should be of the order of the square of consequent linguistic terms that must be differentiated. These conditions are achieved when the number of experiments (observations in this case) goes to infinity. The theoretical results obtained also proved that the results of the Fuzzy Boolean Nets are equivalent to those achieved with non parametric estimation by Parzen windows [9], which implies that every data experiment is supposed to belong to a distribution with a given probability density. This same conjecture is made for proving the relation between the number of inputs per consequent neuron and the granularity. Thus one assumes that there exists a hidden set of qualitative rules relating a set of consecutive data observations (or simple functions of these observations, such as moving averages) to the next unknown value and this set of rules complies with the above assumptions.

The behaviour of the Boolean Fuzzy Net has been studied [13], in terms of the quadratic error obtained when the number of used neurons, of inputs per consequent neuron and per antecedent (mA) and of different rules on the consequent (granularity) are used as parameters. Also the effect of the number of different teaching experiments has been observed. In that study, experimental values for antecedent and consequent have been artificially generated from normal distributed variables, with known average and variances. It has been observed that the error diminishes with the number mA until a level of minimum is achieved, depending on the number of neurons. This optimum value for mA, in those experiments, is of the order of 50 for 100 neurons and of 70 for 1000 neurons, when the value of three is used for consequent granularity. Further increasing of the mA from that point on has the effect of increasing or decreasing the error on an oscillation form. This minimum error, however, depends on the number of consequent neurons and for a granularity of 3 on the consequent (say: "Low", "Medium" and "High") and for a number of the order of 100 experiments this minimum is 0.00063, for 100 neurons and 0.0003 , both for 1000 and 5000 neurons, being the error a value between these two for 500 neurons. These values may be used as a clue for the work of this paper but, due to the different nature of the problem, a more specific investigation has been performed.

In order to evaluate the net performance four different indicators have been calculated as follows. In this context d_p is considered the predicted value of index p and y_p the actual data set value to be predicted. Also, the values of the data set have been standardized for the interval [0,1].

Mean squared error, $MSE = 1/2N\sum_{p}(d_p - y_p)^2$

Normalized Mean Squared error,

$$NMSE = \frac{\sum_p (y_p - d_p)^2}{\sum_p (y_p - \bar{y}_p)^2}$$

Mean Percentage Error,

$$MRPE = 1/N \sum_p \left| d_p - y_p \right| .100\%$$

Percentage of Correct (up/down) Direction Prediction, $PCD = \frac{1}{N} f_p * 100\%$

With $f_p = 1$ if $sign(d_p - y_{p-1}) = sign(y_p - y_{p-1})$; $f_p = 0$ otherwise.

Different teaching and testing sets have been used, namely part of the extra data set (about 9000 points) as teaching set and the main data set (about 1000 points) as the test set or 800 points of the main data set as teaching set and the other 200 points for test set, with similar results.

A set of different values for the parameters have been tried, namely:

R(number of consequent rules) has been changed from 3 to 9, A(the number of antecedents) on the range from 3 to 5, N(number of consequent neurons) has been tried with values 500 and 1000. The number of inputs per antecedent and per consequent neuron, mA, was used according the theoretical conditions stated above, that is, at least of the order of the square of the number of rules. The actual values were from 24 for 3 rules to 161 for 9 rules. Figure 1 displays 200 points both of real and predicted data, of the 1000 used for test.

These values compare optimistically with those already obtained by other authors. In particular one of the best reported results, using Hidden Markov Experts [15], accounts for NMSE values (the other indicators are not available) from 0.01 (using 25 Experts) to 0.139 (using 10 Experts). It should, however, be stressed that that work has used 10 antecedents (comparing with 3 and 9 of this work) and that the so-called Experts have a similar role to Qualitative Rules. Thus, the value of 0.095 (10 antecedents and 20 Experts) for NMSE should be compared to a FBN result using 10 antecedents and 20 rules! A similar result is obtained here with 3 antecedents and 9 rules.

Table 1 describes the obtained values for two of the tests.

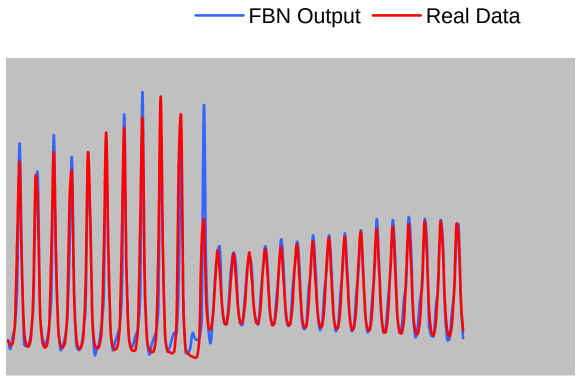

Figure 1: Real and Predicted data

Table 1: Prediction MSE, NMSE, MRPE and PCD

	MSE	NMSE	MRPE	PCD
R=3;A=5; N=500;m A=24	0.0027	0.19	4.2%	88%
R=9;A=3; N=500;m A=161	0.0014	0.099	3.19%	90.5%

This suggests that the interpolation capability of these nets are very well adapted to this type of time series forecast and that the number of past values/antecedents necessary to a good prediction, as long as enough rules are used, is not so high as indicated in other works,.

4 Conclusions

It has been shown that FBN's are a good option for one step ahead prediction for certain time series problems, particularly in non trivial and stationary problems, such as the chaotic pulsations of low dimensionality of the used laser data set. The obtained error rates, both for mean square error and mean relative percentage error, are quite good when compared with other reported results. Also the high percentage of correct direction prediction (in terms of the derivative sign of the series variable) encourages the study the FBN behavior with other time series type where this sign prevision is of most importance, such as stock exchange or currency problems.

Finally, although the main objective of this paper was the forecast of the value of a variable given its passed values, an interesting question is the interpretation of the predictions. That is, the issue of explaining the prediction results by qualitative rules can be very useful for understanding the underlying laws governing the time series behavior. Since the FBN model learns rules this question is automatically answered by listing those rules. Moreover for the test phase it is possible to list also the activation ratio of each rule, instead of exhibiting only the final defuzzified value.

References

[1] G.Box,G.Jenkins and G.C.Reinsel, *Time Series Analysis, Forecasting and Control, Oakland*, California: Holden Day, 1994.

[2] P.Chang and J. Hu, “Optimal nonlinear adaptive prediction and modelling of MPEG video in ATM networks using pipelined recurrent neural networks”, *IEEE J.Select.Areas Commun*., vol 15, Aug. 1997.

[3] S.Chong,S.Li and J.Ghosh, “Predictive dynamic bandwidth allocation for efficient transport of real time VBR over ATM”, *IEEE J.Select. Areas Commun*.,vol13,pp 12-33, Jan. 1995.

[4] A.Geva , “Non-stationary time series prediction using fuzzy clustering”, *Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society*, pp 413-417, 1999.

[5] S.Haykin, *Adaptive Filter Theory*. Upper Saddle River,NJ: Prentice Hall, 1996.

[6] Hebb, D. *The Organization of Behaviour: A Neuropsychological Theory*. John Wiley &Sons. (1949).

[7] U. Huebner,N.B. Abraham and C.Weiss “Dimensions and Entropies of chaotic intensity pulsations in a single-mode far infraredNH3 laser”, *Phys. Rev. A40*, p 6354(1989)

[8] Lin, Chin-Ten and Lee,C.S. *A Neuro-Fuzzy Synergism to Intelligent Systems.New Jersey* : Prentice Hall. (1996).

[9] Parzen, E.(1962). "On Estimation of a probability density function and mode" *Ann. Math. Stat*., 33, 1065-1072.

[10] Tomé,J.A."Neural Activation ratio based Fuzzy Reasoning."*Proc. IEEE World Congress on Computational Inteligence*, Anchorage,May 1998,pp 1217-1222.

[11] Tomé,J.A."Counting Boolean Networks are Universal Approximators" *Proc. of 1998 Conference of NAFIPS*, Florida, August 1998, pp 212-216.

[12] Tomé,J.A. and Carvalho, J.P. "Rule Capacity in Fuzzy Boolean Networks”. *Proc. of NAFIPS-FLINT 2002 International Conference*, New Orleans, IEEE2002, pp 124-128.

[13] Tomé, J.A. “Error convergence on Fuzzy Boolean Nets”. Internal Report, INESC- 2003.

[14] A. Weigeng and N.A. Gershenfeld, Time Series Prediction: Forecasting the Future and Understandibg the Past . Reading,MA: Addison-Wesley, 1999

[15] X.Wang, P. Whigham,Da Deng,M. Purvis, “ “Time-Line” Hidden markov Experts for Time Series Prediction”, *Neural Information Processing – Letters and Reviews*, Vol. 3,2, May 2004.