MEANS FOR ESTIMATING THE FORECASTING ERROR IN FUZZY INDUCTIVE REASONING

François E. Cellier
Electr. & Comp. Engr. Dept.
University of Arizona
Tucson, AZ 85721
U.S.A.
Cellier@ECE.Arizona.Edu

Josefina López
Institut de Cibernètica
Univ. Politècnica de Catalunya
Diagonal 647, 2na. planta
Barcelona 08028, Spain
Lopez@IC.UPC.Es

Angela Nebot
Llenguatges i Sistemes Informàtics
Univ. Politècnica de Catalunya
Diagonal 647, 8na. planta
Barcelona 08028, Spain
Angela@LSI.UPC.Es

Gabriela Cembrano
Institut de Cibernètica
Univ. Politècnica de Catalunya
Diagonal 647, 2na. planta
Barcelona 08028, Spain
Cembrano@IC.UPC.Es

Abstract

This paper deals with the assessment of the reliability of predictions made in the context of the fuzzy inductive reasoning methodology. The reliability of predictions is assessed by means of two different confidence measures, a *proximity measure*, and a *similarity measure*. A time series and a single–input/single–output (SISO) system are used as two different applications to study the viability of these measures.

Keywords: Qualitative Modeling, Qualitative Simulation, Fuzzy Inductive Reasoning, Estimation of Modeling Error.

INTRODUCTION

Models never reflect all facets of reality. Models are always somewhat reductionistic in nature, and consequently, simulation results are never totally reliable. Hence it is important to always interpret simulation results with caution and a certain degree of scepticism.

The degree of uncertainty associated with a model of a system depends heavily on the nature of that system. Simple man–made engineering systems, such as electronic circuits, are characterized by a small degree of uncertainty, since it is an actual design goal when fabricating these systems to keep the degree of uncertainty small. On the other hand, biological or economic systems are usually characterized by a fairly large degree of uncertainty.

Although the request for scepticism is a good mandate on moral grounds, is it also a practical demand? How should, for example, a medical practitioner know how to judge the reliability of a prediction made? He or she has no way of knowing how good the predictions of a simulation model are that he or she may be using. Hence it is important to instil scepticism into the simulation software itself, rather than demanding it of its user.

Assessing the inaccuracy of a simulation result is in itself a modeling task. Yet, the same methodology that is used to model the output to be predicted cannot be used to model its error. This would lead to a paradoxical situation. If it indeed were possible to compute, in a deterministic sense, the inaccuracy of a prediction made, then one could simply subtract the predicted prediction error from the prediction itself and obtain the precise value of the output. Evidently, this cannot be done. The modeling error can only be modeled in a statistical sense.

In this paper, two confidence measures implemented inside the fuzzy inductive reasoning methodology will be described that assess the error of a prediction made simultaneously with making the prediction.

In a robust modeling methodology capable of

dealing with model uncertainty (as qualitative modeling techniques should always be), modeling the modeling error should not be an afterthought. Modeling the output and modeling its error should be done simultaneously. A modeling and simulation methodology that does not take the model uncertainty into consideration from the beginning is not robust when dealing with uncertain situations.

In the next section, the two confidence measures, a *proximity measure* and a *similarity measure*, are described in the context of the Fuzzy Inductive Reasoning (FIR) methodology. A description of the main elements of the FIR methodology can be found in a companion paper published in the same set of proceedings (López 96). A more detailed description of all facets of the methodology can be found in (Cellier 96).

Subsequently, two applications, one related to time series forecasting and the other to SISO systems modeling, are studied in order to discuss the viability and effectiveness of the proposed confidence measures.

CONFIDENCE MEASURES OF THE FIR METHODOLOGY

Before the proximity and similarity measures can be properly presented, it is necessary to introduce some concepts and definitions related to the fuzzy inductive reasoning methodology.

FIR always deals with multi–input/single–output (MISO) systems. Therefore, each state consists of a number of input variables and a single output variable. In the forecasting process, FIR compares the current values of the set of input variables (the so–called input state) with all the input states stored in the experience data base, which was constructed during training, i.e., in the modeling process. It determines, which are the five nearest neighbors in terms of their input states in the experience data base, and estimates the new output value as a weighted sum of the output values of the five nearest neighbors, i.e., proximity is determined in the input space, leading to a set of weight factors that are then used for interpolation in the output space.

In order to create a meaningful metric of proximity in the input space, it is necessary to normalize the input variables. This is accomplished using a so–called pseudo–regeneration (Cellier 91) of the previously fuzzified input variables (López 96):

$$pos_i = class_i + side_i \cdot (1.0 - Memb_i) \qquad (1)$$

where the *class* values are assumed to be integers starting from '1' representing the lowest class, and the *side* values are also integers assuming the values '-1,' '0,' and '$+1$,' representing the logical values '*left*,' '*center*,' and '*right*' of the fuzzy membership function (Cellier 96). The index i represents the i^{th} input variable in the input state of the current observation. The position value, pos_i can be viewed as a normalized pseudo–regeneration of the i^{th} input variable. Irrespective of the original values of the input variable, pos_i assumes values in the range $[1.0, 1.5]$ for the lowest class, $[1.5, 2.5]$ for the next higher class, etc.

Similarly,

$$pos_{ij} = class_{ij} + side_{ij} \cdot (1.0 - Memb_{ij}) \qquad (2)$$

represents the normalized pseudo–regeneration of the i^{th} input variable of the j^{th} nearest neighbor in the experience data base.

$$\mathbf{pos} = [\, pos_1,\ pos_2,\ \ldots,\ pos_n \,] \qquad (3)$$

is the position vector representing the current input state, assuming that the system to be modeled contains n m–inputs (López 96), and:

$$\mathbf{pos_j} = [\, pos_{1j},\ pos_{2j},\ \ldots,\ pos_{nj} \,] \qquad (4)$$

represents the corresponding position vector of the j^{th} nearest neighbor.

The position vectors of the five nearest neighbors are the starting point for computing both types of confidence measures.

The Proximity Measure

The idea behind assessing the reliability by means of the proximity measure is directly related to the distance of the observations available in the experience data base to the new observation. If the new observation is very close to one or several previous observations, then the confidence in the prediction made is high. On the other hand, if the new observation is quite far away from all of its five nearest

neighbors and a lot of interpolation has to be done to determine the new output value, then the prediction should be assessed as less reliable.

The distance between the current input state and its j^{th} nearest neighbor is computed as:

$$dis_j = \|\mathbf{pos} - \mathbf{pos_j}\| \tag{5}$$

In order to prevent a possible division by zero in the proposed algorithm, it is necessary to avoid distance values of 0.0:

$$d_j = \max(dis_j, \epsilon) \tag{6}$$

where ϵ is the smallest number that can be distinguished from 1.0 in addition.

$$s_d = \sum_{j=1}^{5} d_j \tag{7}$$

is the sum of the distances of the five nearest neighbors, and:

$$d_{\text{rel}_j} = \frac{d_j}{s_d} \tag{8}$$

are the relative distances. By applying this algorithm either to the entire experience data base or a suitable subset thereof, the five nearest neighbors can be determined while simultaneously computing their distance functions.

The interpolation is done in the output space. Absolute weights are computed as:

$$w_{\text{abs}_j} = \frac{1.0}{d_{\text{rel}_j}} \tag{9}$$

and

$$s_w = \sum_{j=1}^{5} w_{\text{abs}_j} \tag{10}$$

is the sum of the absolute weights. Hence the relative weights can be computed as:

$$w_{\text{rel}_j} = \frac{w_{\text{abs}_j}}{s_w} \tag{11}$$

The average distance used to determine the confidence measure is computed as a weighted sum of the relative distances of the five nearest neighbors:

$$d_{\text{conf}} = \sum_{j=1}^{5} w_{\text{rel}_j} \cdot d_j \tag{12}$$

The largest possible value of the average distance can be calculated as:

$$d_{\text{conf}_{\max}} = \sqrt{\sum_{i=1}^{n} (n_{\text{cl}_i} - 1)^2} \tag{13}$$

where n_{cl_i} is the number of classes used in the fuzzification of the i^{th} input variable.

Finally, the confidence is evaluated as:

$$conf_1 = 1.0 - \frac{d_{\text{conf}}}{d_{\text{conf}_{\max}}} \tag{14}$$

where *conf* is a *quality measure* (Cellier 91), i.e., a real–valued number in the range $[0.0, 1.0]$. Values close to 1.0 denote a high confidence in the prediction made, whereas values close to 0.0 denote a low confidence in the forecast.

The Similarity Measure

Measures of confidence can also be defined without the explicit use of a distance function. The similarity measure proposed in this paper is a generalization of the classical set–theoretic equality functions. These generalizations rely on the definitions of *cardinality* and *difference* in fuzzy set theory.

The similarity measure presented in this section, which was originally proposed by Dubois and Pradé (Dubois 80) is based on intersection, union, and cardinality:

$$S_1(A, B) = \frac{\| A \cap B \|}{\| A \cup B \|}$$

Clearly, when $A = B$, then $S_1(A, B) = 1.0$, and when A and B are totally disjoint, then $S_1(A, B) = 0.0$.

In FIR, this concept is implemented in the following way. The position variables pos_i assume

values in the range $[1.0, ncl_i]$. They are normalized once more:

$$P_i = \frac{(pos_i - 1)}{(ncl_i - 1)} \tag{15}$$

The P_i variables assume values in the range $[0.0, 1.0]$. Similarly, the re–normalized position value for the i^{th} input variable of the j^{th} nearest neighbor in the normalized experience data base can be computed as:

$$P_{ij} = \frac{(pos_{ij} - 1)}{(ncl_i - 1)} \tag{16}$$

The similarity of the i^{th} input variable of the j^{th} nearest neighbor based on intersection is then defined as follows:

$$S_{ij} = \frac{\min(P_i, P_{ij})}{\max(P_i, P_{ij})} \tag{17}$$

The overall similarity of the j^{th} neighbor is defined as the average similarity of all its input variables in the input space:

$$sim_j = \frac{1}{n} \sum_{i=1}^{n} S_{ij} \tag{18}$$

The similarity measure for the new forecast value is consequently defined as the weighted average of the similarities of the five nearest neighbors:

$$conf_2 = \sum_{i=1}^{n} w_{\text{rel}_j} \cdot sim_j \tag{19}$$

Also $conf_2$ is a quality measure in the range $[0.0, 1.0]$. Values close to 1.0 denote a reliable forecast.

APPLICATIONS

Central Nervous System

In this section, the two previously explained confidence measures, the proximity measure and the similarity measure, are studied in the context of a SISO system, one facet of the cardiovascular system of the human body.

The cardiovascular system is composed of the hemodynamical system and the Central Nervous System (CNS) control. The CNS comprises, among others, the signals that are transmitted from the brain to the heart and to the blood vessels for controlling the hemodynamical system.

A mixed quantitative and qualitative model of the cardiovascular system using FIR to describe the qualitative subsystems has been presented in (Nebot 96). It contains five separate FIR controller models. One of the five controllers that compose the CNS, the *Peripheric Resistance* (PR) controller, is used, in this paper, as an example to study the validity of the two confidence measures presented in the previous sections when applied to systems with input and output signals.

The input of the system is the Carotid Sinus Pressure and the output is the Peripheric Resistance control signal. The *peripheric resistance controller* FIR model is presented in Eq. 20. It is an optimal mask (López 96) of depth five. A set of 5000 data values has been used in the identification process in order to obtain the optimal mask that best captures the behavior of the given system.

$$\begin{array}{c} {}_t\backslash^x \\ t - 4\delta t \\ t - 3\delta t \\ t - 2\delta t \\ t - \delta t \\ t \end{array} \begin{array}{c} \begin{array}{cc} CSP & PR \end{array} \\ \begin{pmatrix} 0 & -1 \\ 0 & 0 \\ -2 & -3 \\ 0 & 0 \\ 0 & +1 \end{pmatrix} \end{array} \tag{20}$$

The model was validated by using it to forecast six different data sets that had not been employed in the model identification process, i.e., using data that the model had never seen before. Each one of these six data sets, with a size of about 600 data points each, contains signals representing specific morphologies, allowing the validation of the model for different system behaviors. The upper portion of Figure 1 shows a comparison of the output obtained by forecasting one of the data sets using the FIR model with the true measured output.

There is only a short interval around the sample 300 where the FIR model was unable to predict accurately how the signal is supposed to continue. This insecurity is related to non–stationary behavioral characteristics of the measurement data. The peripheric resistance control level during various Valsalva maneuvers recorded in the experience data base was always slightly different during

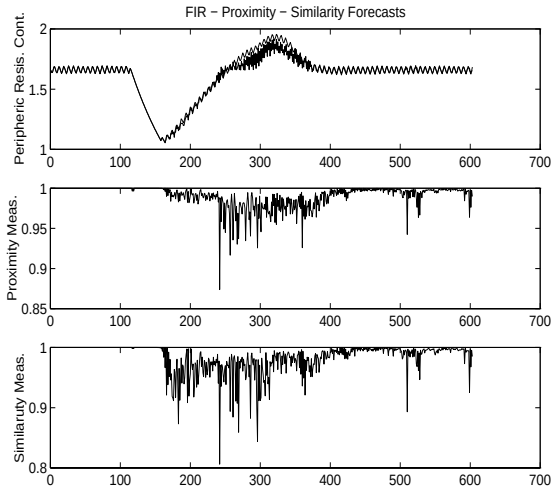

Figure 1: Proximity and Similarity Forecasts for the FIR Peripheric Resistance Controller Model.

this period, and consequently, FIR is insecure as to what precisely it should predict, and oscillates between the different plausible predictions. The forecast shown in Figure 1 represents the poorest result obtained for any of the six testing data sets.

During the qualitative simulation process, the confidence measures are computed together with the forecast. The lower portions of Figure 1 show the two confidence measures, i.e., the *proximity* and the *similarity* measure.

The forecast depicted in the upper portion of Figure 1 shows that the prediction is excellent during the early part of the simulation. It is also quite good during the late part of the simulation period. However, there is a time segment, approximately between samples 170 and 360, where the quality of the prediction is reduced. Between samples 170 and 240, the high–frequency components of the signal are not properly represented, and between samples 240 and 360, the prediction is outright wrong. Both confidence measures respond reliably to the prediction error, as can be seen in Figures 1. The confidence values in the early and late segments of the simulation are very high, whereas they are much reduced in the middle section. It can also be noticed that the similarity measure is more sensitive to the prediction errors than the proximity measure.

Water Demand Time Series

A FIR model has been obtained to predict the daily water demand of a section of the city of Barcelona (López 96). The available measurement data contain the daily water demand of approximately two years. The demand is measured in m^3.

In a first experiment, which is the experiment described in detail in (López 96), a look ahead of one day was performed, i.e., a single new data point was predicted each day on the basis of the true measurements from the previous two weeks. 570 days (from January 1985 to 24 July 1986) were used as training data, whereas the final 128 days (from 25 July to 29 November 1986) were used as testing data. Figure 2 shows the forecasts (dashed line) together with the measured values (continuous line). Underneath, the two confidence measures are depicted.

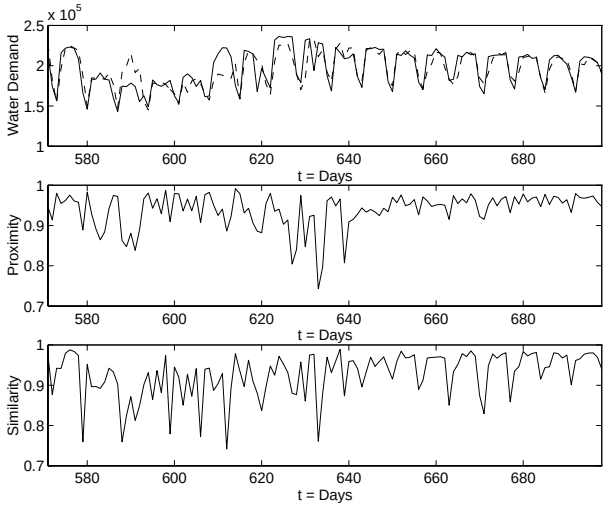

Figure 2: Water Demand FIR single–step prediction from 25.06.1986 - 29.11.1986 together with confidence measures

Contrary to the cardiology example, the relationship between the prediction error and the confidence measures is not immediately evident. The confidence seems to go down during the weekends, which is an artifact of data deprivation more than anything else. There were simply a smaller number of neighbors available in the experience data base representing previous weekend days.

However, the relationship between the prediction error and the two confidence measures can be shown statistically. To this end, the crosscorrelation between the prediction error and $(1.0 - conf_i)$ was computed using Matlab's *xcov* function. The results are shown in Figure 3.

It can be seen clearly that there exists a positive correlation between the two quantities at the

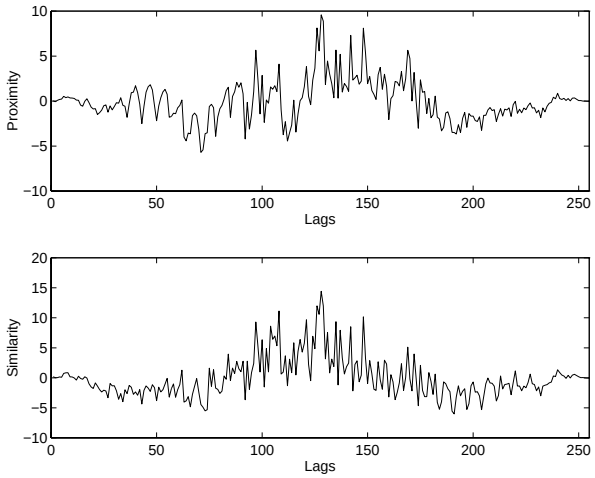

Figure 3: Crosscorrelation between prediction error and $(1 - conf_i)$

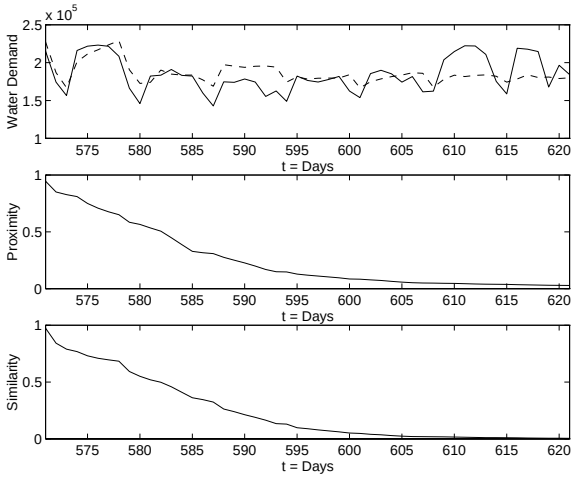

Figure 4: Water Demand FIR multiple–step prediction from 22.11.1986 - 30.11.1986 together with cumulative confidence measures

center, which is what was to be expected if the confidence measure operates correctly. It can also be noticed that the correlation is a little higher in the case of the similarity measure, which is an indication of the somewhat higher sensitivity and reliability of this confidence measure.

In a second experiment, a prediction over multiple steps was performed. As before, the first 570 days were used for training, but then, a prediction was made with a look ahead over 50 days without incorporating any additional measurement data. Consequently, the later forecasts rely entirely on previously made forecasts, and eventually, the quality of the forecast must deteriorate.

Underneath the forecast, the cumulative confidences are given. The cumulative confidence of a point is defined recursively as the product of the local confidence of that point with the cumulative confidence of the previous point. It turns out that a prediction with a look ahead of one week can be done reliably. The second week's prediction is still quite good. The 7–day cycle is preserved for a few more weeks, before even it gets washed away. When the cumulative confidence decays to a value below roughly 50%, the prediction must be considered unreliable, and consequently, SAPS–II, our implementation of FIR, can be instructed to reject making any predictions with too low a cumulative confidence value.

Discussion

The two examples presented above are very useful in analyzing the characteristics of the proposed confidence measures, as well as the capabilities of FIR for prediction in two very different situations. The central nervous system data correspond to a process that is largely deterministic, except at the peaks when the precise value is different from one period to the next, whereas the water demand time series corresponds to a stochastic quasi–stationary process.

The confidence measures proposed above are related to how deterministic the data base is, i.e., how close or disperse the outputs are for any one input pattern. When a process is mostly deterministic, SAPS–II will produce a good model and in most cases, there exists only one possible output for a given input pattern, or possible alternative outputs are very close, so that the confidence will be high and the error will be low. In the few cases where a number of disperse outputs exists for a certain input pattern, it is more likely that a larger prediction error occurs, and this corresponds to a lower confidence zone. This can be clearly observed in the central nervous system example, where the correlation between the prediction error and $(1 - confidence)$ for both measures, is very high. The same is not true, however, in the case of the water demand time series. In this example, the prediction error and $(1 - confidence)$ are still statistically related, at least in the single–step prediction, as shown in Figure 3. The correlation coefficient is 0.3890 in the case of the similarity measure, and 0.3342 in the case of the proximity measure.

CONCLUSIONS

When using fuzzy inductive reasoning models in prediction, it is very important to generate not only forecasts for the output variables, but also measures of the reliability of each forecast. In this context, two measures of the reliability of FIR predictions have been proposed, a *proximity measure* and a *similarity measure*. After testing these measures on a largely deterministic input–output system and on a mostly stochastic time series, a few conclusions can be drawn:

- The similarity measure is more sensitive to the prediction error than the proximity measure. This is reasonable, because the similarity measure preserves more information about the qualitative difference between a new input state and its neighbors in the experience data base than the proximity measure.
- Since the models derived by SAPS are largely deterministic and autoregressive, in both the deterministic and the autoregressive stochastic processes, the proposed measures are useful tools to evaluate the likelihood of errors. More specifically, large proximity or similarity values indicate that a low prediction error is likely to occur.
- In time series corresponding to stochastic processes that are not entirely autoregressive, i.e., processes where the errors may be correlated, there is not necessarily a significant correlation between the prediction error and $(1 - \mathit{confidence})$. Therefore, the correlation between these two entities may, in general, be used as an indicator of how well the series in question may be fitted by an autoregressive or deterministic model.

A remark of a more philosophical nature is in place as well. The better the modeling methodology works, the less likely it is that a measure of the quality of the prediction can be made. If indeed the model were to exploit *all* the information that is available in the measurement data, then the model of the prediction error would necessarily have to behave like uncorrelated white noise, because whatever can be said about the prediction error can, at least in theory, be exploited to improve the model. In practice, this is not a big problem. As long as the prediction error does not behave like white noise, the information obtained is useful to assess the quality of the prediction. On the other hand, once the prediction error starts to behave like white noise, the researcher can be assured that he or she has exploited every bit of knowledge available and has come up with the best possible model already. Hence, even in that case, the error analysis tells us something of value.

The authors are continuing their research work in this area, specifically on the improvement of the modeling and prediction capabilities of SAPS to general stochastic processes, and on the derivation of new confidence measures.

REFERENCES

Cellier, F.E., A. Nebot, F. Mugica and A. de Albornoz.
(1996), "Combined Qualitative/Quantitative Simulation Models of Continuous–Time Processes Using Fuzzy Inductive Reasoning Techniques," *Intl. J. Gen. Syst.*, **24**(1–2), pp. 95–116.

Cellier, F.E. (1991), *Continuous System Modeling*, Springer–Verlag, New York.

Dubois, D. and H. Pradé. (1980), *Fuzzy Sets and Systems, Theory and Applications*, Academic Press, New York.

López, J., G. Cembrano, and F.E. Cellier. (1996), "Time Series Prediction Using Fuzzy Inductive Reasoning," *ESM'96: European Simulation Multiconference*, Budapest, Hungary, June 2-6.

Nebot, A. (1996), *Qualitative Modeling of Biomedical Systems*, Springer–Verlag, New York, Series in Biomathematics, accepted for publication.