Improving the Forecasting Capability of Fuzzy Inductive Reasoning by Means of Dynamic Mask Allocation

Josefina López
Institut d'Informàtica i Robòtica Industrial
Universitat Politècnica de Catalunya
Edifici Nexus
Gran Capità 2-4
Barcelona 08034, Spain
Phone: +34(93)401-5755
Fax: +34(93)401-5750
JLopez@IRI.UPC.ES

François E. Cellier
Electrical & Computer Engineering Dept.
University of Arizona
P.O.Box 210104
Tucson, AZ 85721-0104
U.S.A.
Phone: +1(520)621-6192
Fax: +1(520)621-8076
Cellier@ECE.Arizona.Edu

ABSTRACT

This paper deals with a new extension of the Fuzzy Inductive Reasoning (FIR) methodology that makes use of the estimate of the prediction error generated by FIR for automated dynamic model selection during the simulation run. FIR can choose among several models available in the model library, and dynamically selects the model that is currently most appropriate for the task of making inductively future predictions of system behavior on the basis of observed earlier behavior[1].

Keywords: Fuzzy Systems, Inductive Reasoning, Dynamic Model Allocation.

INTRODUCTION

At the last ICQFN conference in Budapest, it was shown that FIR [Cellier *et al.*, 1996a; Escobet *et al.*, 1999] can make statistically significant estimations of its own prediction error [Cellier *et al.*, 1996b]. In fact, this intrinsic feature may easily be FIR's most important characteristic [López, 1999]. However, it was not attempted to come up with an estimate of the true prediction error directly. Instead, an indirect assessment was obtained in the form of a *confidence measure*.

It can be argued that a direct attempt at estimating the prediction error must be futile as long as FIR does its job, because if it were possible to estimate the prediction error directly, then this estimate could be subtracted from the prediction, leading to an improved prediction. Such a naïve scheme can obviously never work, as long as FIR exploits all of the information available in the measurement data for making its predictions.

However, *any* estimate of the prediction error, even an indirect one, can in principle, be used to improve the accuracy of a prediction made. After all, such an estimate *does* provide additional information about the prediction, an information that should be exploitable. This paper presents one approach to exploiting this information for improving the quality of predictions made.

The same approach can also be used to tackle yet another problem, namely that of dealing with *variable structure systems*. Some systems are *time–varying*. They change their behavioral patterns over time.

Many such systems operate in a number of different predefined *regimes*, i.e., during some period of time, they exhibit similar behavioral patterns, and then, they suddenly switch from one operational mode to another. A car may serve as an example. It is in first gear during some period of time. Suddenly, the driver (or an automatic controller) decides to shift into second gear. The car now behaves differently from before.

Other systems are truly time–variant. They exhibit a continuous range of operational patterns. Here, an approach that classifies the behavioral patterns into discrete regimes is only an approximation of the true system complexity, yet, it may still be an effective way of enabling a person to make predictions of such a system.

In this paper, it will be shown that FIR, together

[1]The research presented in this paper was supported by the TAP-96-0882 CICYT project

with the proposed methodology of *dynamic mask allocation*, can be used to deal with any and all of the above scenarios in a robust fashion.

DYNAMIC MASK ALLOCATION

The idea behind dynamic mask allocation is straightforward. In [Cellier *et al.*, 1996a], it was shown that FIR, in its *qualitative modeling module*, proposes an *optimal mask*, i.e., a set of m–inputs that best characterize the output to be predicted.

Two separate quality measures were used to determine the optimal mask: the entropy reduction measure, H_r, that effectively measures the quality of information available, and an observation ratio measure, OR, that determines the quantity of information available. The mask quality was then determined as the product of the entropy reduction measure and the observation ratio measure:

$$Q = H_r \cdot OR \tag{1}$$

The optimal mask is the one that exhibits the largest Q value.

Yet, the selection of the optimal mask is by no means unique. There usually exist many masks of quite similar mask qualities (with similar Q values). Any of these masks can be used to make decent predictions. In fact, the *foptmask* routine of SAPS–II, the current implementation of FIR, returns not only the optimal mask, but the best mask of each complexity, in order to give the user a choice. Furthermore, a *mask evaluation report* can be requested that lists each mask that was tried together with its Q value.

It is quite reasonable to make multiple predictions in parallel using different masks of high quality. Until now, this was never done, because the user had no means to judge, which of the predictions obtained is the best. Using either of the two confidence measures introduced in [Cellier *et al.*, 1996b], this is now possible. Each of the predictions made using different masks comes with its own confidence estimate. It is then reasonable to accept, in each step, the one prediction that exhibits the largest confidence value.

Figure 1 demonstrates the algorithm. The switch at the left symbolizes the process of sampling, i.e., the passing of time. At each time step, n different FIR models (different masks) are used to make predictions in parallel. The variable y_i represents the predicted

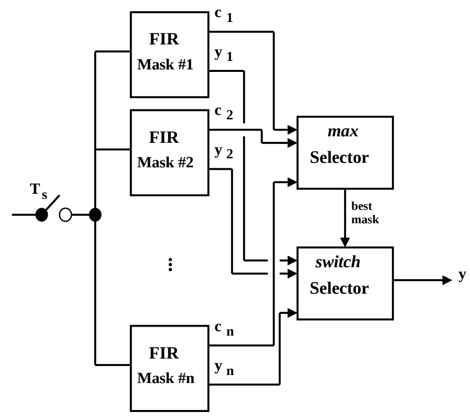

Figure 1: Dynamic Mask Allocation

output using mask m_i, and c_i represents the estimated confidence in the prediction made. The different confidence values are then passed on to a *max selector* that determines the index, i, of the currently best mask:

$$i = \text{index of } \{c_i\}, \quad c_i = max! \tag{2}$$

The predicted outputs are fed into a *switch selector* that also receives the index of the currently best mask from the max selector. The switch selector passes through the y_i associated with the selected c_i:

$$y = y_i \tag{3}$$

The mask allocation is dynamic, because in each step, a different mask may be selected.

DMAFIR AND QDMAFIR

The algorithm explained in the previous section has been named *DMAFIR*, denoting *Dynamic Mask Allocation for FIR*. The algorithm does not take into account the relative quality of the selected mask.

It might make sense to punish the use of masks of lower quality. To this end, a new quality measure is introduced:

$$Q_{\text{rel}_i} = \frac{Q_i}{Q_{\text{opt}}} \tag{4}$$

where Q_i is the mask quality of the selected mask, m_i, and Q_{opt} is the mask quality of the optimal mask. Clearly, Q_{rel_i} qualifies as a quality measure, since the value of Q_{rel_i} is in the range $[0.0, 1.0]$ with a larger value denoting the selection of a higher–quality mask. Q_{rel_i} is a *static mask quality* measure, as neither Q_i nor Q_{opt} change their values over time for any given mask, m_i.

Using this quality measure, the *dynamic mask quality* can be defined as:

$$Q_{\text{dyn}}(t) = Q_{\text{rel}_i}(t) \cdot conf_{\text{sim}}(t) \qquad (5)$$

where $conf_{\text{sim}}(t)$ is the confidence value using the similarity measure at time t. Here, $Q_{\text{rel}_i}(t)$ is indeed a function of time, because during each step, a different mask, m_i, may be chosen.

The so modified algorithm has been named *QDMAFIR*, denoting *Quality–adjusted Dynamic Mask Allocation for FIR*.

In the sequel, the two algorithms, DMAFIR and QDMAFIR, shall be applied to the water demand series of the city of Barcelona [López *et al.*, 1996] to check whether dynamic mask allocation might help in obtaining better predictions. The rationale behind this experiment is that the water demand is quite different during weekends than during work days. Thus, if a mask is offered to FIR that makes better predictions for holidays, and another mask is provided that makes better predictions for working days, then FIR might automatically and dynamically choose the best mask in each case, offering overall better predictions than either of the individual masks might be able to generate.

BARCELONA WATER DEMAND PREDICTION

In [López *et al.*, 1996], a time series had been introduced that represents the water demand of the City of Barcelona. It had been shown that FIR can successfully predict the future behavior of this series.

Unfortunately, there are not enough data points available to train a model that predicts particularly well during weekends. Hence it was decided to offer to DMAFIR and QDMAFIR the top masks of complexities 2, up to 8, as proposed by the *mhis* matrix of the *foptmask* routine of SAPS–II. These masks, together with their qualities, are listed in Table 1.

The best mask is a mask of complexity 4. It uses the values one day back, one week back, and two weeks back for its prediction. This is reasonable because of the weekly cyclic behavior of the time series. The second best mask is a mask of complexity 5. The masks of yet higher complexity offer a considerably lower quality, because the amount of available data does not justify their use.

Table 1: Suboptimal Masks and Their Qualities for Barcelona Time Series

Mask	Mask Qual.
$y = \tilde{f}(y(t-\delta t), y(t-7\delta t), y(t-14\delta t))$	0.4539
$y = \tilde{f}(y(t-\delta t), y(t-3\delta t), y(t-7\delta t), y(t-12\delta t))$	0.3997
$y = \tilde{f}(y(t-\delta t), y(t-7\delta t))$	0.3879
$y = \tilde{f}(y(t-7\delta t))$	0.2993
$y = \tilde{f}(y(t-\delta t), y(t-3\delta t), y(t-5\delta t), y(t-11\delta t), y(t-14\delta t))$	0.2280
$y = \tilde{f}(y(t-\delta t), y(t-3\delta t), y(t-5\delta t), y(t-7\delta t), y(t-11\delta t)), y(t-14\delta t))$	0.0988
$y = \tilde{f}(y(t-\delta t), y(t-3\delta t), y(t-5\delta t), y(t-7\delta t), y(t-11\delta t)), y(t-13\delta t), y(t-14\delta t))$	0.0374

The mask quality is a compromise measure between two competing components. The *entropy reduction* measure, H_r, assesses the uncertainty associated with a prediction, i.e., it is a measure of the *quality of information* available. The *observation ratio* measure, OR, judges the quality of neighbors, i.e., it is a measure of the *quantity of information* available.

Because of the lack of available training data, FIR cannot justify to always use a mask of high complexity. However, if at any point in time, there happen to be good neighbors available, then a mask of high complexity may offer a higher local quality, because it is associated with less uncertainty.

Both DMAFIR and QDMAFIR allow to exploit this.

At any point in time, FIR will look for the proximity (or similarity) of its nearest neighbors, and it will pick the mask of highest complexity that offers neighbors that are sufficiently close.

Figure 2 compares the prediction errors of FIR when using only the optimal mask with that of FIR using the DMAFIR algorithm, once with the similarity confidence measure, and once with the proximity confidence measure [Cellier *et al.*, 1996b].

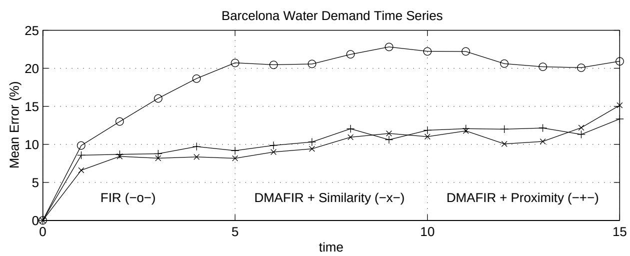

Figure 2: Comparison of FIR and DMAFIR for Barcelona Time Series

The independent axis shows the number of days in the future for which a prediction is made, whereas the dependent axis shows the average prediction error observed for that day using a formula with four separate components. The first two punish deviations in mean and standard deviation, whereas the other two components punish the absolute and dissimilarity errors between the normalized curves, i.e., after normalizing the mean of both curves to zero and their standard deviation to one [López, 1999].

There is a dramatic reduction in prediction errors. The proximity and similarity measures [Cellier *et al.*, 1996b] offer similar error reductions, with the similarity measure being slightly better on average.

Figure 3 compares the prediction errors of FIR when using only the optimal mask with that of FIR using the QDMAFIR algorithm, once with the similarity measure, and once with the proximity measure.

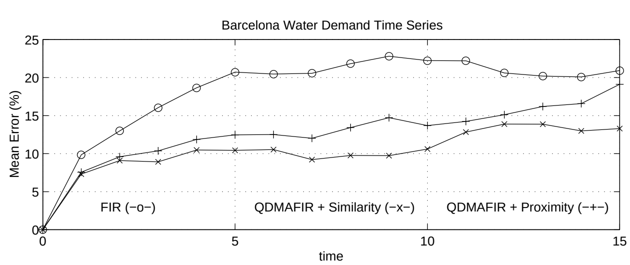

Figure 3: Comparison of FIR and QDMAFIR for Barcelona Time Series

The results are quite similar to those found above. However in this case, the similarity measure offers a consistently larger error reduction than the proximity measure.

From now on, only the similarity measure will be used, because it was shown experimentally to be the better overall measure of the two.

Figure 4 compares the prediction errors of FIR when using only the optimal mask with that of FIR using the DMAFIR and QDMAFIR algorithms together with the similarity measure.

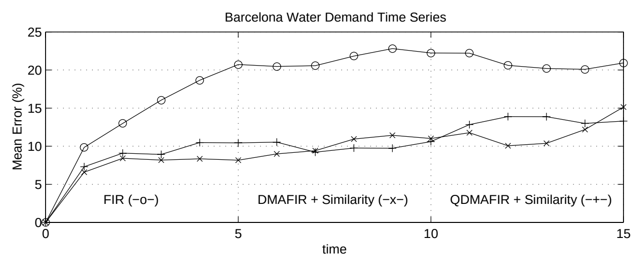

Figure 4: Comparison of FIR, DMAFIR, and QDMAFIR for Barcelona Time Series

It turns out that the DMAFIR algorithm offers better results than the QDMAFIR algorithm. This is understandable. QDMAFIR gives a preference to masks that are close to the optimal mask in complexity. This hampers the ability of the algorithm to pick the mask of highest complexity that locally offers good neighbors.

The improvement of the forecasting quality obtainable by using a dynamic mask allocation algorithm is quite remarkable. Hence the fact that FIR offers a self–assessment capability is pivotal to its success in making predictions about the future behavior of time series. Prediction methods that do not offer a self–assessment capability are therefore severely disadvantaged.

MULTIPLE REGIMES

In this section, it will be demonstrated that the DMAFIR algorithm can be used to predict time series that operate in multiple regimes, i.e., where the behavioral patterns change between time segments. To this end, a new time series is introduced: the Van–der–Pol oscillator series.

The Van–der–Pol oscillator is described by the following second–order differential equation:

$$\ddot{x} - \mu \cdot (1 - x^2) \cdot \dot{x} + x = 0 \qquad (6)$$

By choosing the outputs of the two integrators as two state variables:

$$\begin{aligned} \xi_1 &= x \\ \xi_2 &= \dot{x} \end{aligned} \tag{7}$$

the following state–space model is obtained:

$$\begin{aligned} \dot{\xi_1} &= \xi_2 \\ \dot{\xi_2} &= \mu \cdot (1 - \xi_1^2) \cdot \xi_2 - \xi_1 \\ y &= \xi_2 \end{aligned} \tag{8}$$

The ξ_2 variable is used as output of the time series.

This is a synthetic time series, generated by a simulation model. Therefore, the data set can be made as long as needed. The Van–der–Pol oscillator is characterized by a stable limit cycle, i.e., already after the transitory period that is caused by the initial conditions imposed on the model has died down, a single limit cycle (one period of the oscillation) will suffice to characterize the time series completely. The series is thus as active as it can ever be.

The behavioral patterns of the series depend on the choice of the parameter μ. A time series operating in multiple regimes can be created by toggling between different values of μ in the course of the simulation.

To start the experiment, three different models were identified using three different values of μ, namely $\mu = 1.5$, $\mu = 2.5$, and $\mu = 3.5$. The first 80 data points of each time series were discarded, as they represent the transitory period. The next 800 data points were used to learn the behavior of each series, and the subsequent 200 data points were used as testing data. With a sampling rate of 0.05, 200 data points correspond roughly to one oscillation period, i.e., four limit cycles were used for training the model, and one limit cycle was used for testing. The mask depth was chosen to be 50. All variables were classified into five classes with the landmarks -7.0, -0.5, -0.25, $+0.25$, $+0.5$, and $+7.0$. The same landmarks were used for all three time series, such that the results of the predictions can be more easily compared with each other.

The models obtained in this way are shown in Table 2.

The mask qualities are very high because of the strictly deterministic nature of the series. The optimal masks for $\mu = 2.5$ and $\mu = 3.5$ are identical, yet the input/output behaviors will be different because of the different training data used by the two models.

Table 2: Optimal Masks and their Qualities for Van–der–Pol Series

Regime	Optimal Mask	Mask Qual.
$\mu = 1.5$	$y = \tilde{f}(y(t-\delta t), y(t-47\delta t))$	0.9342
$\mu = 2.5$	$y = \tilde{f}(y(t-\delta t))$	0.9085
$\mu = 3.5$	$y = \tilde{f}(y(t-\delta t))$	0.9146

Figure 5 compares the true time series with their predictions for each of the three models.

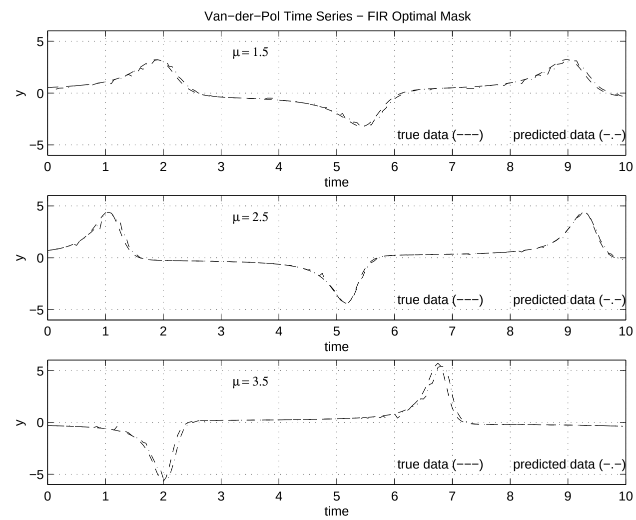

Figure 5: One–day Predictions of the Van–der–Pol Series Using FIR Without Dynamic Mask Allocation

The top graph in Figure 5 compares the true Van–der–Pol cycle for $\mu = 1.5$ with the FIR predictions obtained using the model obtained for the same series. The graph below compares the Van–der–Pol data for $\mu = 2.5$ with the FIR predictions obtained using the corresponding FIR model, etc.

Because of the completely deterministic nature of this time series, the predictions should be perfect. They are not perfect due to data deprivation. Since 800 data points were used for training, the experience data base contains only *four* cycles. Thus, when FIR, during the prediction, looks for *five* good neighbors, it only encounters *four* that are truly pertinent.

Figure 6 shows the predictions obtained when applying the model (optimal mask plus training data) obtained for the time series with $\mu = 1.5$ to the other two time series.

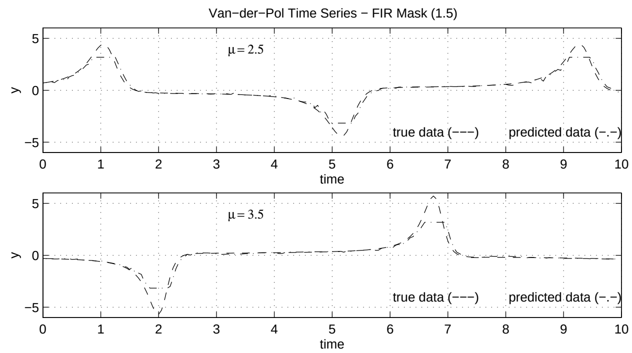

Figure 6: One–day Predictions of the Van–der–Pol Series Using FIR With $\mu = 1.5$ Model

The model cannot predict the peaks of the time series with $\mu = 2.5$ and $\mu = 3.5$ correctly, because it has never seen such tall peaks. FIR can only predict behaviors that it has seen before.

Similar results are obtained when using the other two time series with the wrong models. Table 3 summarizes the errors obtained.

Table 3: Prediction Errors for Van–der–Pol Series

Series	$\mu = 1.5$	$\mu = 2.5$	$\mu = 3.5$
Model ($\mu = 1.5$)	2.6292	6.7597	10.3922
Model ($\mu = 2.5$)	2.9645	0.9747	4.6463
Model ($\mu = 3.5$)	4.2691	2.5744	1.8272

The results are as they would have been expected. The values along the diagonal are smallest, and the values in the two remaining corners are largest. It also makes sense that the model obtained for $\mu = 3.5$ is more capable of predicting the series with $\mu = 1.5$ than the other way around.

Next, a time series shall be constructed, in which the variable μ assumes a value of 1.5 during one segment, followed by a value of 2.5 during the second time segment, followed by yet another time segment, in which $\mu = 3.5$. The multiple regimes series consists of 553 samples.

Figure 7 shows the results of predicting the multiple regimes series using the three models independently.

The model obtained for $\mu = 1.5$ cannot predict the higher peaks of the second and third time segment very well, therefore its error must be largest. The model

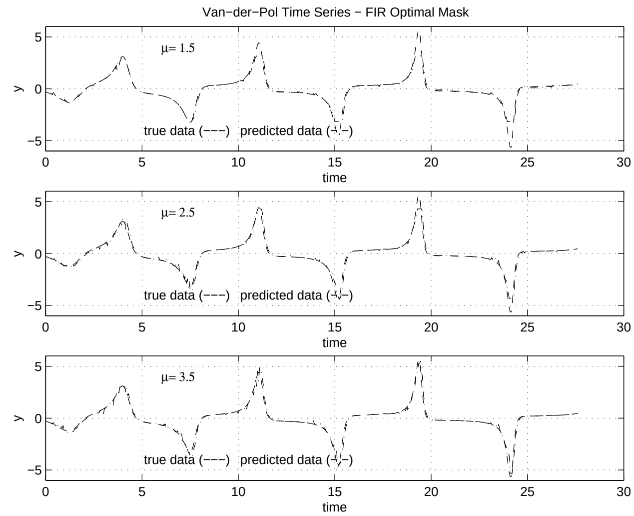

Figure 7: One–day Predictions of the Van–der–Pol Multiple Regimes Series

obtained for $\mu = 3.5$ does a decent job at predicting all three segments. Thus, its error must be smallest.

Figure 8 shows the results of predicting the multiple regimes series using DMAFIR together with the similarity confidence measure. The three individual models (optimal masks plus training data sets) are offered to the DMAFIR algorithm to choose from.

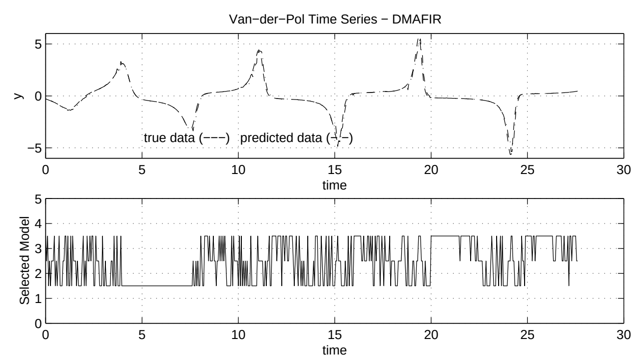

Figure 8: One–day Predictions of the Van–der–Pol Multiple Regimes Series Using DMAFIR

The top plot shows the prediction obtained by DMAFIR. The bottom plot shows, which of the three models was chosen at any point in time. The value plotted is the μ–value of the chosen model. During the first time segment, consisting of the first 178 samples, the "average" μ–value is $\mu_{\text{avg}} = 1.9831$. During the second segment, the average μ–value is $\mu_{\text{avg}} = 2.4831$. Finally, during the third time segment, the average μ–value is $\mu_{\text{avg}} = 3.0871$. Thus, on average, FIR indeed picks more often than not the correct model.

Table 4 lists the prediction errors obtained for the different simulations using the modified error formula.

Table 4: Prediction Errors for Multiple Regimes Van–der–Pol Series

	error
Model for $\mu = 1.5$	5.8759
Model for $\mu = 2.5$	2.2978
Model for $\mu = 3.5$	1.9317
DMAFIR	1.1195

As was to be expected, the model obtained for $\mu = 3.5$ shows the smallest of the individual errors. However, the error obtained using DMAFIR is still considerably smaller. This demonstrates that DMAFIR can indeed be successfully applied to the problem of predicting time series that operate in multiple regimes.

VARIABLE STRUCTURE SYSTEMS

In this section, it will be shown that the DMAFIR algorithm can be successfully employed for predicting time–varying systems. Whereas a system that operates in multiple regimes exhibits a fixed number of different behavioral patterns, a time–varying system exhibits an entire spectrum of different behavioral patterns.

To demonstrate DMAFIR's ability of dealing with time–varying systems, the Van–der–Pol oscillator is used once again. This time, a series was generated, in which μ changes its value continuously in the range $[1.0, 3.5]$. The time series contains 953 records sampled using a sampling interval of 0.05. The value of μ changes once per sample.

Figure 9 shows the results of predicting the time–varying series using the three models independently.

Each peak is of slightly different amplitude, i.e., the time–varying Van–der–Pol oscillator series is no longer completely deterministic. As expected, the model obtained for $\mu = 3.5$ works best, because it has no difficulty predicting the high–amplitude peaks. Also the model obtained for $\mu = 2.5$ works very well, because the system has low–pass characteristics. Although μ varies in the range $[1.0, 3.5]$, the extremely small and extremely large peaks characteristic of very small and very large μ values never show up in the simulation results. The model obtained for $\mu = 1.5$ is least suitable, because it cannot predict high–amplitude peaks that it has never seen during the training phase.

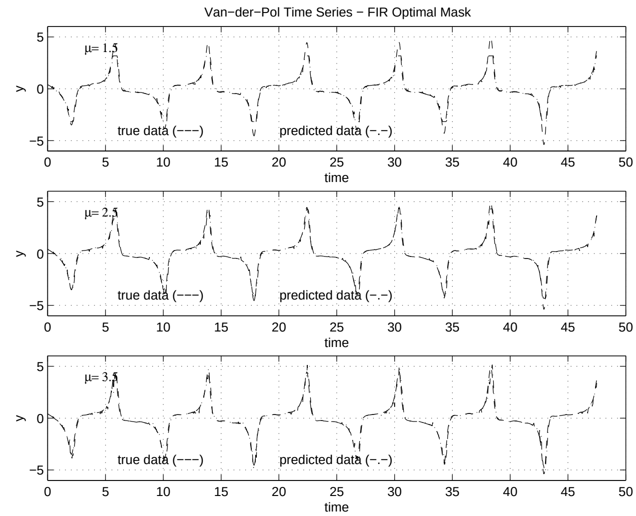

Figure 9: One–day Predictions of the Van–der–Pol Time–Varying Series

Figure 10 shows the results of predicting the time–varying Van–der–Pol series using DMAFIR together with the similarity confidence measure. The three individual models (optimal masks plus training data sets) were offered to the DMAFIR algorithm to choose from.

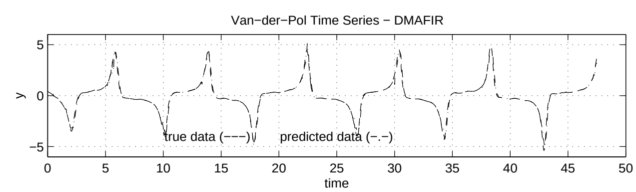

Figure 10: One–day Predictions of the Van–der–Pol Time–Varying Series Using DMAFIR

The prediction is close to perfect. As expected, DMAFIR makes the prediction more robust, and reduces the prediction error to a level that is below that obtainable by either of the individual models.

Table 5 lists the prediction errors obtained for the different simulations using the modified error formula.

Table 5: Prediction Errors for Time–Varying Van–der–Pol Series

	error
Model for $\mu = 1.5$	5.7431
Model for $\mu = 2.5$	1.4864
Model for $\mu = 3.5$	1.8791
DMAFIR	1.2997

The experiment shows that DMAFIR is indeed capable of dealing with variable structure system predictions. Although such systems do not have a finite set of individual behavioral patterns, it is useful to discretize the spectrum of behavioral patterns, identify individual models for each of these patterns, and then let DMAFIR choose among them during the variable structure system prediction.

CONCLUSIONS

In this paper, a methodology was introduced that allows to exploit the confidence measure of FIR, an indirect prediction error estimate, for improving the predictions made.

Since a direct error estimate coupled with an error subtraction scheme does not work, confidence measures had been introduced in [Cellier *et al.*, 1996b] as a means to indirectly assess the quality of predictions made. The present paper has shown how this indirect information can be exploited to improve the quality of the predictions made by FIR. It was shown that the self–assessment capability of FIR is pivotal to its capability of making high–quality predictions of time series.

REFERENCES

Cellier F.E., A. Nebot, F. Mugica and A. De Albornoz, Combined qualitative/quantitative simulation models of continuous–time processes using fuzzy inductive reasoning techniques, Intl. J. General Systems 24 (1–2), pp. 95–116, 1996a.

Cellier, F.E., J. Lṕez, A. Nebot, and G. Cembrano, Means for Estimating the Forecasting Error in Fuzzy Inductive Reasoning, Proc. ESM'96, European Simulation MultiConference, Budapest, Hungary, pp. 654–660, 1996b.

Escobet, T., A. Nebot, and F.E. Cellier, Model Acceptability Measure for the Identification of Failures in Qualitative Fault Monitoring Systems, Proc. ESM'99, European Simulation MultiConference, Warsaw, Poland, 1999.

López, J., Time Series Prediction Using Inductive Reasoning Techniques, Ph.D. dissertation, Universitat Politècnica de Catalunya, 1999.

López, J., G. Cembrano, and F.E. Cellier, Time Series Prediction Using Fuzzy Inductive Reasoning: A Case Study, Proc. ESM'96, European Simulation MultiConference, Budapest, Hungary, pp. 765–770, 1996.