

Multi-resolution Time–Series Prediction Using Fuzzy Inductive Reasoning

François E. Cellier
Department of Electrical and
Computer Engineering
University of Arizona
Tucson, AZ 85721-0104
U.S.A.
E-mail: cellier@ece.arizona.edu

Àngela Nebot
Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Jordi Girona Salgado, 1-3
Barcelona 08034
Spain
E-mail: angela@lsi.upc.es

Abstract—The paper describes a new approach to multi-resolution prediction of time series using Fuzzy Inductive Reasoning (FIR). The time series is decomposed into a trend series and another series describing the deviation from the trend. The two time series are then predicted independently of each other, and the two predictions are superposed in the end. The trend series is obtained by means of a moving average, whereas the deviation series is obtained by a process of de-trending using “daily return” calculations. The paper deals both with interpolation and with extrapolation problems.

I. INTRODUCTION

Multi-resolution time series are series that contain both low-frequency and high-frequency components. They are difficult to predict, because the high-frequency component of the series forces the prediction algorithm to use small time steps, whereas the low-frequency component of the series calls for a large time horizon. Thus, the prediction must be carried across many steps, in which prediction errors may be accumulated.

The CATS time series used in this paper is shown in Fig.1.

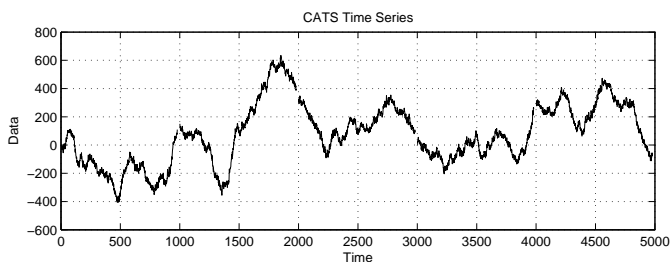


Fig. 1. CATS time series

The CATS time series is a synthetic time series, consisting of 5000 data records containing five gaps of 20 records each. The data values from 981–1000, 1981–2000, 2981–3000, 3981–4000, and 4981–5000 are missing. The CATS time series was proposed as a prediction contest for the 2004 IJCNN Conference.

The series is a multi-resolution series as it exhibits a very irregular high-frequency oscillation superposed over a fairly irregular low-frequency oscillation. The aim is to fill the five

gaps with predicted values, predicting simultaneously the low-frequency component and the high-frequency component of the series. The first four of the gaps represent an interpolation problem, whereas the final gap represents an extrapolation problem.

II. THE PREDICTION PROBLEM

Most of the more advanced prediction techniques can be subdivided into two separate tasks. In a first step, the *modeling step*, the algorithm uses a set of training data to identify a model of a process, from which the training data could have been obtained. In a second step, the *simulation step*, the algorithm uses the previously identified model to make predictions outside the training data set.

The modeling algorithm can either attempt to identify the true structure of the system, from which the training data were obtained, or it can content itself with identifying any process able to explain the training data set. In the former case, we talk about a *deep model*, whereas models in the latter category are referred to as *shallow models*.

The identified model can be either a *quantitative* or a *qualitative* model. A quantitative model operates on the measurement data directly, whereas a qualitative model first discretizes the measurement data, and then reasons about the discrete classes only.

The model can be either a *parametric model* or a *non-parametric model*. A parametric model maps the knowledge contained in the training data set onto a set of model parameters. During the simulation phase, the training data are no longer needed, since the information contained in them is now stored in the parameter values. A non-parametric model only classifies the training data during the modeling phase, and refers back to these classified training data during the simulation phase.

III. FUZZY INDUCTIVE REASONING

The Fuzzy Inductive Reasoning (FIR) methodology offers a model-based approach to predicting either univariate or multivariate time series [1], [2]. A FIR model is a qualitative, non-parametric, shallow model based on fuzzy logic.

In a first step, the available measurement data are fuzzified. Thereby, the real-valued quantitative data values are mapped onto qualitative triples, consisting of a class value, a fuzzy membership value, and a side value. The process is illustrated in Fig.2 by means of a variable, called ambient temperature.

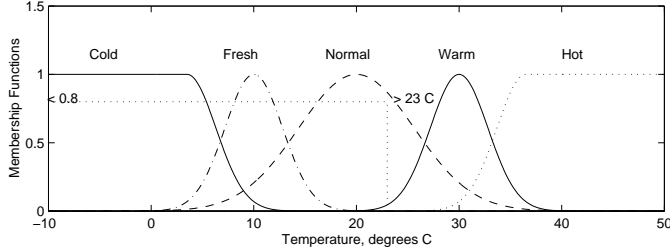


Fig. 2. Fuzzification in FIR

The ambient temperature is mapped onto five discrete classes, called *cold*, *fresh*, *normal*, *warm*, and *hot*. Each of these classes is associated with its own fuzzy membership function, a function with values in the range $[0.0, 1.0]$. The fuzzy membership functions can either be Gaussian distributions or triangular distributions. The side function can assume values of either *left*, *center*, or *right*. Within each fuzzy class, values to the left of the maximum of the fuzzy membership function have associated a side value of *left*, etc. Hence, an ambient temperature of 23 degrees Centigrade is classified as *normal*, with a fuzzy membership value of 0.8, and a side value of *right*.

Most dialects of fuzzy logic associate qualitative doubles rather than triples with each quantitative value. They ignore the side value, yet allow multiple doubles to represent the same quantitative value. Hence an ambient temperature of 23 degrees Centigrade would be classified as $(normal, 0.8)$ and as $(warm, 0.05)$. The FIR dialect of fuzzy logic uses fuzzy triples, yet only records the most likely triple, i.e., the one associated with the largest fuzzy membership value.

No information is being lost in the process of fuzzification. The qualitative triple can be mapped unambiguously onto a single quantitative value by means of de-fuzzification.

The FIR modeling engine reasons only about the class values. In the case of a univariate time series, the next value of the variable, $x(t + \Delta t)$, must be a function of previous recordings of that same variable:

$$x(t + \Delta t) = f(x(t), x(t - \Delta t), x(t - 2\Delta t), \dots) \quad (1)$$

The FIR modeling engine does not try to identify the function, f . It only determines, which subset of previous recordings is most useful in determining the next value of the variable, x , e.g.

$$x(t + \Delta t) = f(x(t - 5\Delta t), x(t - 2\Delta t), x(t)) \quad (2)$$

which would be represented as a so-called *optimal mask*:

$$mask = [-1, 0, 0, -2, 0, -3, +1] \quad (3)$$

where the +1 element denotes the position of the output within the time series, whereas the negative values in the mask denote the relative positions of the three inputs. In the above example, the 1st input is six time steps in the past relative to the output.

The FIR modeling engine searches through all possible masks up to a given mask depth, creating for each mask an input/output table of class values. The optimal mask is the one that makes the map from the set of input classes to the single output class as deterministic as possible. The FIR modeling engine optimizes the information content of the map by minimizing the Shannon entropy measure.

The FIR modeling engine allows the specification of missing data values, as required for the application at hand [3].

Once the optimal mask has been found, FIR stores the training data for later retrieval in an *experience data base* consisting of an alpha-numerically sorted list of input/output data, whereby each quantitative input/output data record is converted into a record of qualitative triples.

The FIR simulation engine predicts values of the output variable beyond the end of an episode of recorded data values. It uses the previously found optimal mask. The inputs to the mask are inside the known episode, i.e., have known values. They are fuzzified, and a qualitative input record is created that can be compared with the records in the experience data base. The five nearest neighbors are retrieved, and the output value is predicted as a qualitative triple representing a weighted average of the output values of the five nearest neighbors in the experience data base.

If multiple steps are to be predicted, the newly found value is added to the episode of known values, the mask is shifted one data point to the right, and the process is repeated, until the gap has been filled.

IV. LOW-FREQUENCY PREDICTION

In order to produce a low-frequency prediction, a process of smoothing was applied to the original time series. To this end, a moving average was computed for every 10th data point consisting of the mean value of the data point itself and its nine nearest neighbors to the left and to the right.

The low-frequency time series consists of 500 data points. It still contains five gaps, but each gap is now only three data points wide. The low-frequency time series is depicted in Fig.3.

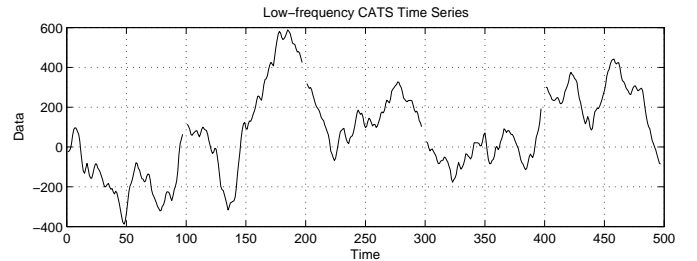


Fig. 3. Low-frequency CATS time series

A FIR model was identified for the low-frequency CATS time series, and the missing values in the five gaps were

predicted using time steps of 1 unit in the low-frequency time series, corresponding to time steps of 10 units in the original time series.

The interpolation problem was handled a little differently from the extrapolation problem. Rather than predicting only three values for each gap, we predicted four values. The fourth value ought to coincide with the first measurement value of the next episode. If it wasn't, we applied a linear correction to the predicted values in the gap to make that value coincide with the next measurement data point. Of course, in the case of the last gap, we could not apply this correction, since we were dealing with an extrapolation problem.

V. HIGH-FREQUENCY PREDICTION

We could have applied the same technique to the original data set, thereby producing directly the high-frequency prediction. We tried this, but the results were not very good. The trend of the curve seemed to be in the way. There were not enough nearest neighbors in the experience data base to come up with good high-frequency predictions.

Thus, we used a de-trending technique, called the "daily return" method. The name comes from the stock market, where this technique is frequently used. The daily return of a time series is another time series, defined as:

$$y(t) = \frac{x(t) - x(t - \Delta t)}{x(t)} \quad (4)$$

The new time series, $y(t)$, is quasi-stationary. Since the original time series had both positive and negative values, we first normalized the original time series to a mean value of 3.0, and a standard deviation of 1.0, using the formula:

$$\xi(t) = \frac{x(t) - \mu}{\sigma} + 3.0 \quad (5)$$

where μ denotes the mean value of the original time series, and σ denotes its standard deviation. The new mean value of 3.0 was chosen such that the normalized time series should have values in the range [1.0, 5.0]. We then applied the daily return formula to the normalized time series, $\xi(t)$. Fig.4 depicts the high-frequency time series, $y(t)$.

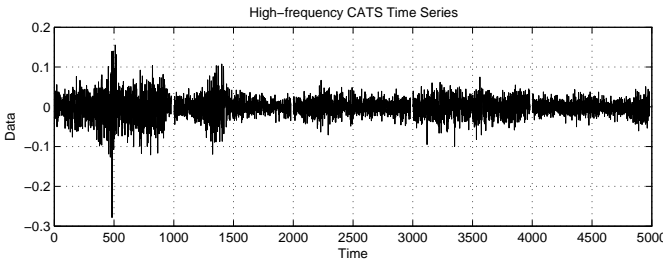


Fig. 4. High-frequency CATS time series

Unfortunately, the high-frequency time series looks almost like white noise. It doesn't contain much information, thus we shouldn't expect to get too much out of it, as white noise is exceedingly hard to predict! Yet, we should hopefully at least be able to get the frequency right.

The high-frequency time series contains six gaps. The first data point of the series is missing, then there are four gaps with a width of 21 points each, and a final gap with a width of 20 points.

We created the optimal FIR model. It was a model of depth 17 with 6 inputs in different locations. We then used that model to predict the first 10 values of each gap, e.g. from point 981–990. We then undid the transformations:

$$\xi(t + \Delta t) = \frac{\xi(t)}{1.0 - y(t + \Delta t)} \quad (6)$$

and:

$$x(t) = (\xi(t) - 3.0) \cdot \sigma + \mu \quad (7)$$

This provided a forecast of the points 981–990 of the original time series. We then treated the value at point 990 from the low-frequency prediction as a measurement data point, and applied a linear correction to get the predicted value from the high-frequency prediction to coincide with that of the low-frequency prediction. We repeated this for all five gaps.

We then de-trended the augmented time series once more, and calculated another high-frequency prediction, using the same model and even the same experience data base, but a different immediate past to predict another 10 data points for each of the five gaps. Since we already had a predicted value for time 5000, each of the sub-gaps could now be treated as an interpolation problem.

VI. RESULTS

The next three figures show the vicinity of the five gaps, after the predicted values have been filled in.

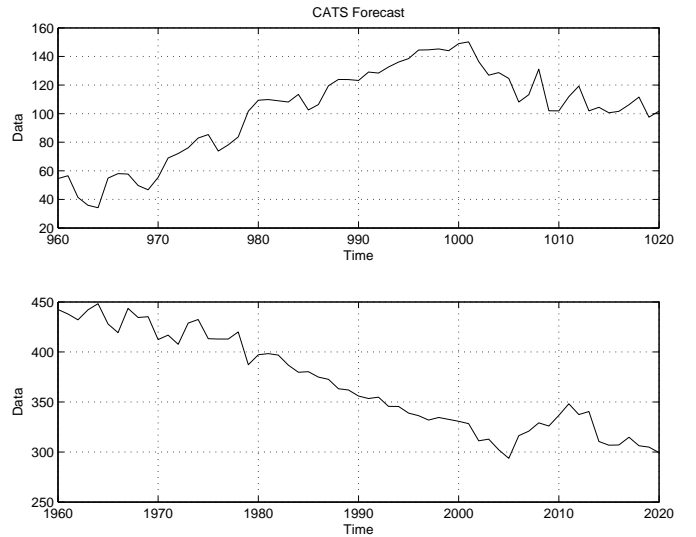


Fig. 5. Predictions of gaps 1 and 2

It seems that the high-frequency prediction worked in so far as getting the frequency of the high-frequency oscillation about right. The amplitudes may be a bit flat. Interesting is the last prediction. It looks like the curve corrected itself upward between points 4970–4980. Yet, the low-frequency model predicted another decline thereafter.

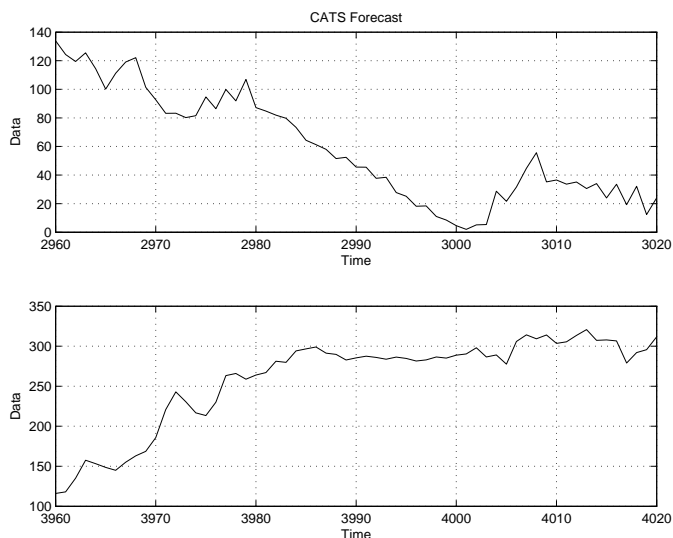


Fig. 6. Predictions of gaps 3 and 4

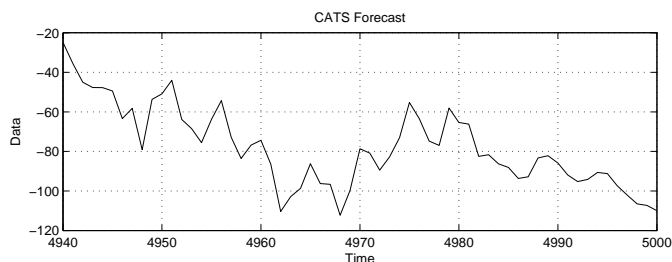


Fig. 7. Prediction of gap 5

VII. VALIDATION

To validate the models, we created other gaps of width 20, for which we already knew the correct values, and applied our methodology to make predictions. Evidently, the predictions should be better than the trivial prediction of connecting the last known value before the gap with the first known value after the gap by a straight line. On average, our predictions were slightly better, but unfortunately, not by much.

The problems are formidable. The low-frequency model is too irregular with too few data points to create high-quality predictions, and the high-frequency model is essentially looking at white noise.

For this reason, we are doubtful that any model can do much better than ours, except possibly by an accidental hit. Highly sophisticated prediction techniques, such as FIR, work very well if the data are fairly regular. They can effectively and efficiently identify nonlinearities, and produce high-quality predictions that a simpler technique could not even dream to produce.

Yet, the more irregular the data are, the better will simple techniques fare by comparison. For the given time series, straight-line approximations are almost as good as predictions using sophisticated models.

VIII. CONCLUSIONS

In this paper, we have presented a new two-layered approach to forecasting multi-resolution time series by decomposing the time series into a smooth low-frequency series and a detrended high-frequency series that are then predicted independently of each other.

The individual predictions were made using two separate FIR models. However, the two-layered approach to multi-resolution forecasting is essentially independent of the underlying forecasting algorithm. The same approach could also be combined with any other forecasting technique.

Why did we choose FIR as the underlying forecasting algorithm? The reason was not that FIR is necessarily the best technique to predict the time series at hand. The time series is so irregular that a simpler prediction algorithm may work just as well. Our reason for using FIR was simply to be able to compare FIR, which is a methodology that we developed over the past 20 years, with other approaches that other researchers may have come up with, in particular with neural network-based approaches.

Our experience has been that FIR works well, when feed-forward neural networks work well, and vice-versa. FIR has some advantages and some disadvantages over feed-forward neural networks. On the one hand, its predictions are a bit more reliable, because it refers back to the training data during simulation. Thus, FIR is incapable of making wild predictions. On the other hand, FIR may not be able to generalize as well as a neural network could, and for precisely the same reasons.

REFERENCES

- [1] F. E. Cellier, 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, 1996.
- [2] J. López, *Time Series Prediction Using Inductive Reasoning Techniques*, Ph.D. Dissertation, Universitat Politècnica de Catalunya, Barcelona, Spain, 1999.
- [3] A. Nebot and F. E. Cellier, "Dealing With Incomplete Data Records in Qualitative Modeling and Simulation of Biomedical Systems," *Proc. CISS'94, 1st Joint Conf. of Intl. Simulation Societies*, Zurich, Switzerland, pp. 605–610, 1994.