

A Flexible Fuzzy Inductive Reasoning Approach for Load Modelling Able to Cope With Missing Data

Sergio Jurado

Sensing&Control Systems S.L.
Aragó 208-210
08011 Barcelona, Spain
+34 605 565 303

sergio.jurado@sensingcontrol.com,
s.juradogomez@gmail.com

Àngela Nebot

Soft Computing research group,
Computer Science Department
Technical University of Catalonia
Jordi Girona 1-3
08034 Barcelona, Spain
+34 934137783

angela@lsi.upc.edu

Francisco Mugica

Soft Computing research group,
Computer Science Department
Technical University of Catalonia
Jordi Girona 1-3
08034 Barcelona, Spain
+34 934134087

fmugica@lsi.upc.edu

ABSTRACT

Load forecasting in buildings and homes has been in the last years a task of increasing importance. New services and functionalities can be offered in the home environment due to this predictions, for instance, the detection of potential demand response programs and peaks that may increase the energy bill in a dynamic tariff framework. Almost real-time predictions are key for these services but missing values can dramatically affect the performance of the energy forecasting or distort the prediction significantly. Fuzzy Inductive Reasoning has been proof to model load consumptions with high accuracy compared to other typical AI and statistical techniques. Nevertheless, it has several limitations when missing data is presented in the training data of the model and during the prediction. In this paper, we present an improved version of the Fuzzy Inductive Reasoning, called Flexible FIR Prediction that can cope with missing information in the input pattern as well as, in situations of not found patterns in the behavioural matrix. The new technique has been tested with real data from one building of the Universitat Politècnica de Catalunya (UPC) and the results show that Flexible FIR Prediction is able to generate good predictions with low errors although missing data is present in the training and online prediction phases.

Categories and Subject Descriptors

I.2.3 [Artificial Intelligence]: Deduction and Theorem Proving – Answer/reason extraction, Deduction, Uncertainty, "fuzzy," and probabilistic reasoning.

General Terms

Algorithms, Performance, Reliability.

Keywords

Prediction with Missing Values, Fuzzy Inductive Reasoning, Energy Modelling.

1. INTRODUCTION

In the last years the development of the Smart Grid is growing fast, encouraged by the environmental targets and the long-term goals set out by the European Commission (EC) Energy Roadmap [1]. The smart metering rollout sustained by the EU [2] not only in the electricity but also in heating, gas and water is a key component to offer new functionalities and services to citizens. When coupled with smart metering systems, the smart grids reach consumers and suppliers by providing information on real-time consumption. With smart meters, consumers can adapt – in time and volume - their energy usage to different energy prices

throughout the day. They will save money on their energy bills by consuming more energy in lower price periods.

In addition to that, the inclusion of renewable energy production in the medium and low voltage and the good omen for home batteries [3] requires new solutions and technologies to cope with them. In concrete, we focus on the importance of forecasting algorithms and the ability to perform good short-term predictions within a continuous flow of data.

Load forecasting has been in the last years a task of increasing importance specially in the medium voltage management; compute the energetic balances, plan the scheduling and execution of the operation, elaborate the plans of expansion and replacements, calculate margin of expected losses and profits for the agents, etc. [4][5]. However, energy forecasting applications also extends to buildings and homes. New services and functionalities can be offered in the home environment due to these predictions. For instance, the detection of potential demand response programs, peaks that may increase the energy bill in a dynamic tariff framework or use them to create messages to change the energy behaviour of the tenants.

Research in load forecasting has led in the development of more than 1000 journal papers until 2009 [6]. However, with this new horizon where dwellings play a more important role thanks to smart meters, different forecasting difficulties arise, when a bottom-up approach (prediction of low voltage loads) is followed instead of typical high and medium voltage forecasting problems.

One of these challenges is that, sometimes, these predictions cannot cope with missing data. The information that arrives from the different sensors in the home area network and/or the smart meters, have problems that may hinder or even unable the forecasting of the next hours and days. This issue is observed in projects such as iURBAN¹ and GreenCom², where either the smart metering infrastructure or the smart home gateway, occasionally, is not sending the data correctly and/or sending missing values. This may be caused by a loose in the Internet connection of the gateway, fail in the communication between smart meters and concentrators, or issues between the database interfaces.

Missing values can dramatically affect the performance of the energy forecasting or distort the prediction significantly. Most time series models assume that the observations are sampled with the same frequency, but it is common to find missing data. The

¹ www.iurban-project.eu

² www.greencom-project.eu

missing data problem has been dealt from many perspectives, for instance, from a state-space methodology. In [7] and [8] different Kalman filters are used to tune the parameters of an ARIMA model with missing data. Gómez et al. proposed in [10] filling the holes in the missing data with arbitrary values and carrying out the maximum likelihood estimation with additive outliers, also in ARIMA models.

In other researches the focus has been on improving exponential smoothing methods. In those cases, the missing data problem has usually been dealt using the Kalman filter [10][11]. In [12] Bermúdez et al. deal with the missing data using an alternative formulation for Holt's model with additive errors that simplifies both the calculus of maximum likelihood estimators of all the unknowns in the model and the calculus of point forecast.

The problem of the missing data has been faced from an offline perspective. In fact, none of these studies is dealing with the missing data in an online forecasting. Therefore, they may present problems in real-time applications.

In this paper, we address the missing data problem found out in a Soft Computing technique called Fuzzy Inductive Reasoning (FIR) [14][17]. Although its popularity is not comparable to other Artificial Intelligence (AI) techniques such as Random Forest (RF), Artificial Neural Networks (ANN) or Supported Vector Machine (SVM). This methodology has been proved to model load consumptions with high accuracy compared to other typical AI and statistical techniques [14] [20]. Nevertheless, it has several problems when missing data are presented in the input variables of the model.

We present an improved version of the Fuzzy Inductive Reasoning that can cope with missing information in the input pattern as well as, in situations of not found patterns in the behavioural matrix. This robust methodology is going to be integrated in a next generation smart home energy management solution, which is called Local Decision Support System (LDSS) and deployed in two European cities: Plovdiv and Rijeka.

The paper is structure as follows: in Section 2 the standard FIR methodology is summarized, the problem of missing values in this methodology is identified and a solution called Flexible FIR Prediction is presented. In Section 3 the experiments performed and results obtained are presented, which includes a comparative analysis with three different approaches to deal with the missing data problem in FIR. Finally, Section 4 pointed out the conclusions of our research and the near future work.

2. FLEXIBLE FUZZY INDUCTIVE REASONING PREDICTION

2.1 Standard FIR

The conceptualization of the FIR methodology arises of the General System Problem Solving (GSPS) approach proposed by Klir [15]. This methodology of modelling and simulation has the ability to describe systems that cannot be easily described by classical mathematics or statistics, i.e. systems for which the underlying physical laws are not well understood [16]. A FIR model is a qualitative non-parametric model based on fuzzy logic. The FIR model consists of its structure (relevant variables or selected features) and a pattern rule base (a set of input/output relations or history behaviour) that are defined as if-then rules.

The process of obtaining a FIR model structure corresponds to a Feature Selection Process (FSP). The model structure holds the relevant features and it is represented by a mask through which

the causal relations (both spatial and temporal) between input and output variables are described. Table 1 presents an example of mask for a system with four inputs (u_1, u_2, u_3, u_4) and one output (y) variables.

Table 1. Example of mask for a system with four inputs (u_1, u_2, u_3, u_4) and one output (y)

x \ t	u_1	u_2	u_3	u_4	y
$t - 3\delta t$	-1	0	0	0	-2
$t - 2\delta t$	0	-3	0	0	0
$t - \delta t$	0	0	0	-4	0
t	0	0	0	0	+1

Each negative element in the mask exhibits a causal relation with the output, i.e. it influences the output up to a certain degree. The single positive value denotes the output. In the example of Table 1, the prediction of the output at the current time, $y(t)$, is directly related to the variables u_1, u_2, u_4 and y in different times, i.e. $u_1(t-3\delta t)$, $y(t-3\delta t)$, $u_2(t-2\delta t)$ and $u_4(t-\delta t)$.

The optimal mask function of FIR is used to obtain the best mask, i.e. the best FIR structure, for the system under study [17]. The procedure consists in finding the mask that best represents the system by computing a quality measure for all possible masks, and selecting the one with the highest quality. The process starts with the definition of a so-called mask candidate matrix encoding an ensemble of all possible masks from which the best is to be chosen. Then the best from these masks is chosen. Table 2 shows an example of mask candidate matrix for the same system example of Table 1.

The mask candidate matrix contains elements of value -1, where the mask has potential causal relations. Elements of value +1 can also be found, where the mask has its output. Finally, elements of value 0 denote forbidden connections.

Table 2. Example of mask candidate matrix for a system with four inputs (u_1, u_2, u_3, u_4) and one output (y)

x \ t	u_1	u_2	u_3	u_4	y
$t - n\delta t$	-1	-1	-1	-1	-1
...
$t - 2\delta t$	-1	-1	-1	-1	-1
$t - \delta t$	-1	-1	-1	-1	-1
t	-1	-1	-1	-1	+1

The number of rows of the mask candidate matrix is called the depth of the mask. It represents the temporal domain that can influence the output. Each row is delayed relative to its successor by a time interval of δt representing the time lapse between two consecutive samplings. δt may vary from one application to another. In the study presented in this paper, a value of δt equal to 1 hour is used, due to the data characteristics.

The optimal mask function of FIR, offers the possibility to specify an upper limit to the acceptable mask's *complexity*, i.e. the largest number of non-zero elements that the mask may contain. Starting from the candidate matrix with minimum complexity two, i.e. 1 input and the output, the qualitative model identification process looks for the best out of the legal masks. Then it is proceed, by searching through all legal masks of complexity three, i.e. all masks with two inputs and the output, and find the best of those. It continues in the same way until the maximum complexity has been reached. This strategy corresponds to an exhaustive search of exponential complexity. However, suboptimal search strategies of polynomial complexity can also be used, i.e. genetic algorithms [13].

Each of the possible masks is compared to the others with respect to its potential merit. The optimality of the mask is evaluated with respect to the maximization of its forecasting power that is quantified by means of the quality measure. Let us focus on the computation of the quality of a specific mask. The overall quality of a mask, Q_m , is defined as the product of its uncertainty reduction measure, H_r , and its observation ratio, O_r , as described in equation 1.

$$Q_m = H_r \cdot O_r \quad (1)$$

The uncertainty reduction measure is defined in equation 2.

$$H_r = 1 - H_m/H_{max} \quad (2)$$

Where H_m is the overall entropy of the mask and H_{max} the highest possible entropy. H_r is a real number in the range between 0.0 and 1.0, where higher values usually indicate an improved forecasting power. The masks with highest entropy reduction values generate forecasts with the smallest amounts of uncertainty. The highest possible entropy H_{max} is obtained when all probabilities are equal. And zero entropy is encountered for totally deterministic relationships. The overall entropy of the mask is then computed as described in equation 3.

$$H_m = - \sum_{\forall i} p(i) \cdot H_i \quad (3)$$

Where $p(i)$ is the probability of that input state to occur and H_i is the Shannon entropy relative to the i^{th} input state. The Shannon entropy relative to the i^{th} input state is calculated from the equation 4.

$$H_i = \sum_{\forall o} p(o|i) \cdot \log_2 p(o|i) \quad (4)$$

Where $p(o|i)$ is the ‘conditional probability’ of a certain output state o to occur, given that the input state i has already occurred. The term probability is meant in a statistical rather than in a true probabilistic sense. It denotes the quotient of the observed frequency of a particular state in the episodically behaviour divided by the highest possible frequency of that state. The observation ratio, O_r , measures the number of observations for each input state. From a statistical point of view, every state should be observed at least five times [18]. If every legal input state has been observed at least five times, O_r is equal to 1.0. If no input state has been observed at all (no data are available), O_r is equal to 0.0. The optimal mask is the mask with the largest Q_m value, being the one that generates forecasts with the smallest amount of uncertainty, and, therefore, the features that compose the structure of this model are the ones selected as the most relevant ones.

Once the most relevant features are identified they can be used in any modelling methodology. The mask is used to obtain the pattern rule (called behaviour matrix) from the fuzzified training data set. Each pattern rule is obtained by reading out the class values through the ‘holes’ of the mask (the places where the mask has negative values), and it places each class next to each other to compose the rule.

Once the behaviour matrix and the mask are available, a prediction of future output states of the system can take place using the FIR inference engine, as described in Figure 1. This process is called qualitative simulation. The FIR inference engine is based on the K-Nearest Neighbour (KNN) rule, commonly used in the pattern recognition field. The forecast of the output variable is obtained by means of composition of the potential conclusion that results from firing the k rules whose antecedents have best matching with the actual state.

As can be seen in the left hand side of Figure 1, the mask is placed on top of the qualitative data matrix (fuzzified test set), in such a way that the output matches with the first element to be predicted. The values of the inputs are read out from the mask and the behaviour matrix (pattern rule base) is used, as it is explained latter, to determine the future value of the output, which can then be copied back into the qualitative data matrix. The mask is then shifted further down one position to predict the next output value. This process is repeated until all the desired values have been forecast.

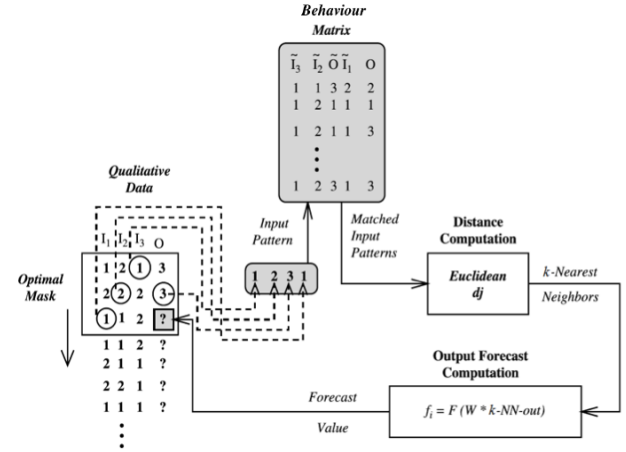


Figure 1. Qualitative simulation process diagram (with an example containing three inputs and one output)

The fuzzy forecasting process works as follows: the input pattern of the new input state is compared with those of all previous recordings of the same input state contained in the behaviour matrix. For this purpose, a normalization function is computed for every element of the new input state and an Euclidean distance formula is used to select the KNN, the ones with smallest distance, which are used to forecast the new output state. The contribution of each neighbour to the estimation of the prediction of the new output state is a function of its proximity. This is expressed by giving a distance weight to each neighbour, as shown in Figure 1. The new output state values can be computed as a weighted sum of the output states of the previously observed five nearest neighbours.

The FIR methodology is, therefore, a modelling and simulation tool that is able to infer the model of the system under study very quickly and it is a good option for real time forecasting. Moreover, it is able to deal with missing data as it has been already proved in a large number of applications [16]. However, its capacity to deal with missing data decreases significantly when the complexity of the mask is big, because it implies the generation of a big number of pattern rules in the behaviour matrix containing missing values. Particularly, in the next Section 2.1 we deal with this problem.

2.2 FIR model with missing data

As explained in detail in the previous section a model in FIR methodology is composed of the mask (model structure) and a set of rules, called pattern rule base (behaviour matrix). The mask defines the causal and temporal relations between the inputs and output variables, i.e. it contains the variables selected as relevant. Once the best mask has been identified, it can be applied to the qualitative data obtained from the system resulting in a particular

pattern rule base, which is a set of rules that represent pseudo-static relationships and that contains the system's behaviour.

This process is illustrated in Figure 2, for one of the subsets with missing values. The left hand side of this figure shows an excerpt of the qualitative data matrix that stores the class values. In this example a class value of -9 indicates a missing value. The dashed box symbolizes the mask that is shifted downwards along the class value matrix. The round 'holes' in the mask denote the positions of the relevant selected features, whereas the square 'hole' indicates the position of the output. The class values are read out from the class value matrix through the 'holes' of the mask, and are placed next to each other in the behaviour matrix that is shown on the right hand side of the figure. Here, each row represents one position of the mask along the class value matrix. Each row of the behaviour matrix represents one pseudo-static qualitative state or pattern rule.

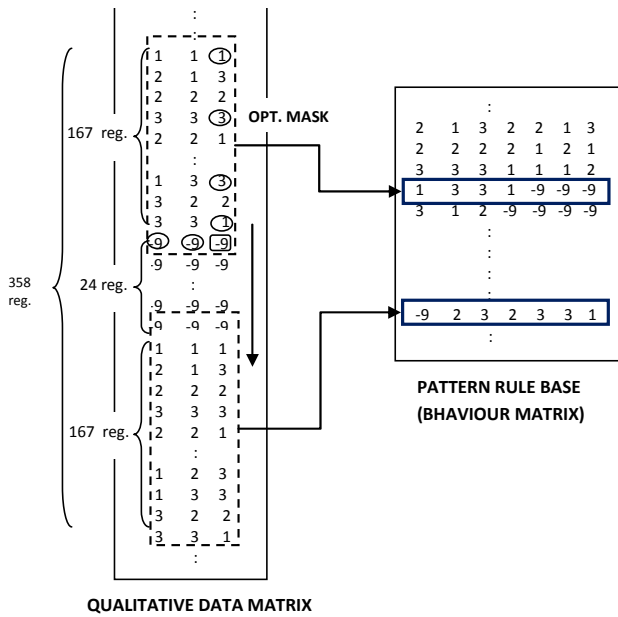


Figure 2. FIR process that generates the pattern rule base starting from the qualitative training data set and the optimal mask. Example that contains consecutive missing values in the input data. The figure illustrates the proportion of generated pattern rules that contains missing values

From the example illustrated in Figure 2, it can be seen that for a qualitative data matrix of 358 registers (reg.) containing 24 consecutive missing values and a mask depth of 168 it is possible to generate up to 191 pattern rules that contain at least one missing element. This can become a huge problem due to the fact that the current prediction process of FIR methodology discards the pattern rules containing missing values, and, therefore, the valid pattern rule base available is reduced significantly. This implies that FIR prediction process, very often, is not able to predict a new input pattern due to the fact that it does not exist in the behaviour matrix.

The current FIR prediction process discards those pattern rules stored in the behaviour matrix that contains one or more missing elements. This means that performs the inference using only the set of pattern rules that has the same class input pattern than the one to which we want to predict the output. Furthermore, the class input patterns of this set are all complete, i.e. do not contain missing values. In consequence, a lot of information is missed.

In addition to this problem, there is a second weakness of FIR when it faces missing values. In concrete, when the input pattern contains missing values. It may happen, especially with online predictions that the input pattern generated after the fuzzification contains missing value. As explained in Section 1 there are different reasons why the information feed into the model may contain missing values: failure in the communication between smart meter and concentrator, the battery of a home area sensor is depleted, loose of internet connection, etc. As it is shown in Figure 3, in the standard version of FIR, the input pattern is searched in the behaviour matrix, however, it cannot find the pattern because, as previously explained, i) the behaviour matrix discard pattern rules containing missing elements and ii) the input pattern could represent several states because it contains a missing value, which could belong to any class of the fuzzified variable.

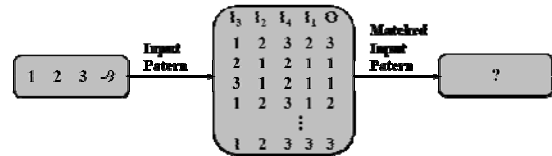


Figure 3. Input pattern containing missing values, which is unable to find a matched in put pattern in the behaviour matrix and therefore unable to predict

2.3 Flexible FIR Prediction

The enhancement proposed in this particular paper is to develop an algorithm that makes the inference process flexible in a dynamic way. The idea is to use the traditional algorithm, when there exist in the behaviour matrix rules that have the same input pattern (free of missing values) than the one to be predicted. When this is not the case, the algorithm will select the set of pattern rules that have the same input pattern but relaxing one of its inputs. That is, the same input pattern but allowing that one of the inputs is missing. Basically, this means that we are using a different mask that the one selected by FIR in the modelling process. Therefore, we are ignoring one of the variables selected as relevant in the feature selection process. It is possible that the "new" mask used is a good suboptimal one, i.e. very close to the quality of the optimal mask of the same complexity. However, it can also happen that the "new" mask to be a bad one.

The new FIR flexible prediction algorithm is depicted in Figure 4. After the training of the model, when a new (online) class input pattern arrives, it is checked if it contains or not missing values. If not, it continues with the standard FIR prediction. On the other hand, if it contains missing values, the new algorithm relax the input pattern, in concrete, the position(s) where there is(are) a missing value(s).

Coming back to the case when the input pattern does not contain missing values, if there is a match of the input pattern in the behaviour matrix, the FIR prediction continues with the standard process. If not, the new algorithm relaxes the input pattern as it will be explained later.

Then, the Maximum Relax Parameter (MRP) is checked. The MRP specifies the maximum number of relaxed parameters of the input pattern. The more relaxed the parameters from the input pattern are, the bigger the search space of forecasting values is. And, therefore, more noise is added to the final prediction. The total number of missing elements allowed in the pattern rules to perform the inference process is determined from the size of the input pattern:

$$MRP = (Size\ Input\ Pattern) / 2 \quad (5)$$

Thus, considering the example of Figure 3, the *Size in Input Pattern* is 4 and the *MRP* is 2, which means that the input pattern will be relaxed up to 2 times. If the maximum relax is reached and there is none match in the behaviour matrix, the forecast value assigned is the same as the previous forecasted value. In hourly energy forecasting problems this solution is usually more convenient than not predicting, because current and previous hourly consumptions are highly correlated.

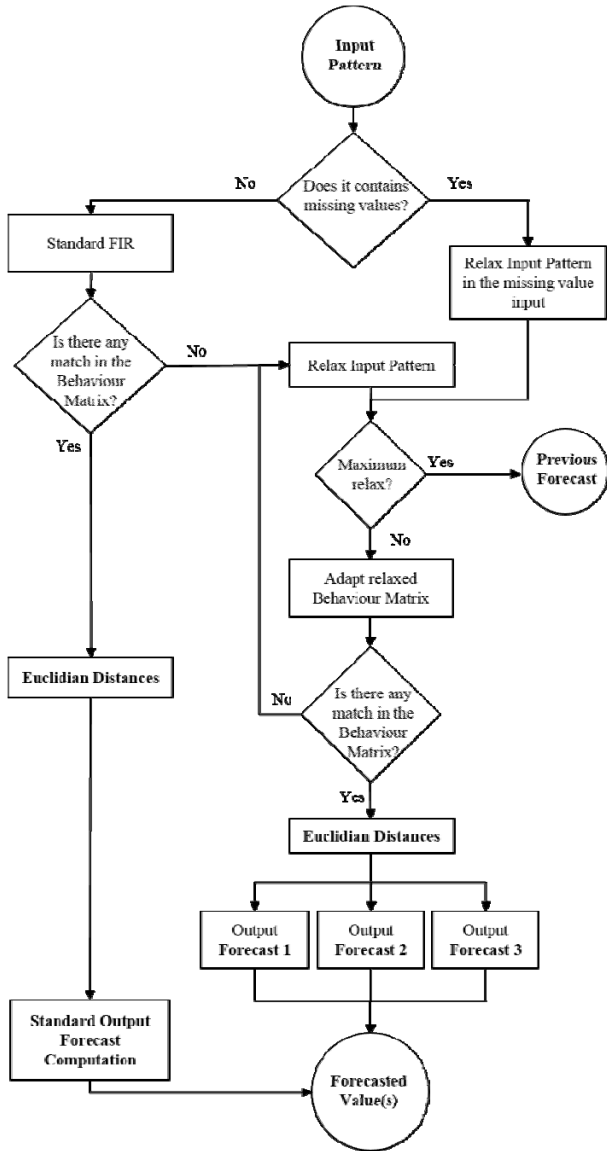


Figure 4. Flow diagram for a more flexible prediction approach in FIR

If the *MRP* is not reached, the behaviour matrix is relaxed to include all registers without the input parameter relaxed. As Figure 5 shows, is the same input pattern but allowing that one of the inputs is missing. It is used a different mask that the one selected by FIR in the modelling process, therefore, ignoring one of the variables selected as relevant in the feature selection process. Similarly, when the input pattern is not found in the behaviour matrix, n Adapted Behaviour Matrix are generated being n all different combinations of relaxation (in the Figure 5

$n=4$) Afterwards, is verified if there is any match of the new relaxed input pattern in the relaxed behaviour matrix.

If it is not possible to find in the behaviour matrix an input pattern with no or only one missing element, the algorithm will look for patterns with two missing elements, penalizing this fact in the inference process. The total number of allowed elements missing in the pattern rules to perform the inference process will be determined in function of the size of the input pattern and the quality of the suboptimal masks associated to the input patterns.

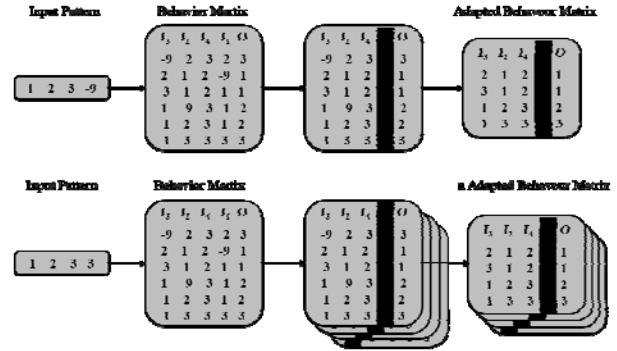


Figure 5. Adaptation of the behaviour matrix when there are missing values in the input pattern (top) or the input pattern is not found in the original behaviour matrix (down)

2.3.1 Output Forecast Computation

As explained in the standard FIR section, the fuzzy forecasting compares the current input pattern with previous recordings that have the same input pattern in the behaviour matrix. Then an Euclidean distance formula is used to select the KNN that are used to forecast the new output state. The contribution of each neighbour to the estimation of the prediction of the new output state is a function of its proximity.

In the new approach, i.e. FIR flexible prediction, this may result in a dispersion of the output classes, due to the loose of information in the input pattern and in the behaviour matrix when they have been relaxed. This would indicate a difficult prediction with low confidence. To cope with this problem, in this research we suggest 3 different *Output Forecast* strategies (In Figure 4 they are represented beneath the computation of the Euclidian Distances):

1. Output Forecast 1: Classic KNN (with K)
2. Output Forecast 2: If first two neighbours have the same distance but they belong to different classes, then, take KNN of the first two classes. If not classic KNN.
3. Output Forecast 3: From all matched input state, take KNN from the most repeated class.

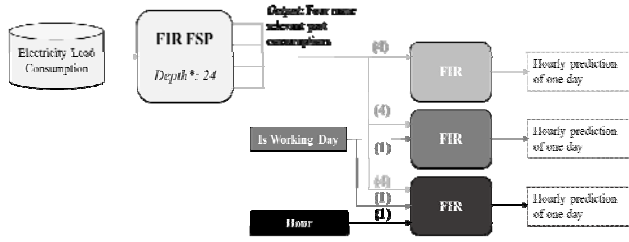
In the following section the different results are presented and discussed.

3. EXPERIMENTS AND RESULTS

3.1 Experiments

The work has been divided in three separated experiments described in Figure 6. Each of the experiments is divided in two stages: the FSP and the model development. Since one of the aims of these experiments is to understand how possible is to perform predictions with partial information in the input pattern, it has

been decided to create three experiments with different number of input variables.



* Same for depth 72 and 24+24

Figure 6. Scheme of the experiment when depth is equal to 24. The same scheme applies to depths 72 and 24+24

The first experiment only takes into account the four most significant historical electricity load consumption values selected in the FIR FSP. The second experiment includes the past values of the electricity load consumption, besides a binary variable that specifies if it is or not a working day. Therefore, the second experiment has five input variables. Finally, the third experiment includes all five variables of the previous experiment plus the hours of the day variable. The FSP has been applied to different depths: 24, 72 and 24 + 24. The reason why these depths are chosen is explained in section 3.1.1.

3.1.1 Model Parameters

In this study the modelling process consists in: 1) feature selection process 2) use of the relevant features to derive a FIR model and 3) selection of the parameters that will allow a more flexible prediction of FIR in case of missing data. This process is repeated for each location and each depth studied. It is decided to build up one model that predicts electricity consumptions one day ahead for each season instead of 4 independent models, i.e. one per season. The main reason is that the data studied do not present a clear trend; hence no deseasonal pre-processing is applied because we want to study the capacity of the different methodologies to obtain generic models.

The FSP is applied only to the historical consumption data and not to the hourly and daily information. It is decided to follow this strategy because the hourly and daily information contains only the hour of the day and if it is or not a working day, respectively. And, therefore, the valuable information is gained with the actual value not with the previous ones. However, previous consumptions contain information patterns from where important knowledge could be extracted.

The selection of the depth and number of variables is a crucial issue that can affect those methods that are more sensitive to the curse of dimensionality. When the FSP of FIR is used, when increasing the number of variables (i.e. complexity) and depth, the quality of the results increases as well until the optimal values are reached. After that, increasing the number of variables may add noise to the system and end up with a result with lower quality. It has been empirically determined that more than four variables and depth higher than 72, do not increase significantly the quality of the FSP of FIR, while computational cost (in terms of time) does exponentially.

Therefore, in order to catch the most relevant previous consumptions in the electric load series, different depths are studied: i) previous 24 hours, ii) previous 72 hours and iii) 24 + 24: previous 24 hours and the past 24 hours of the previous week

(48 past values in total that corresponds to a depth of the mask of 168).

Other additional parameters to set in FIR methodology is the complexity of the masks parameter. In [20] a deep study with different complexities is performed and it concludes that the mask with higher quality and reasonable computational time, was always the mask that selects four past values of the electrical load consumption. Therefore, the four past consumptions values selected by FIR are taken into account in all the models developed in this work and the mask increases when is added the variable is working day and hour.

Regarding the fuzzification parameters three classes and the equal frequency partition algorithm have been used to discretize the electrical load consumption and hour of the day variable. Is working day variable is binary and, therefore, it has been discretized into two classes.

Finally, in the output forecast the parameter set for the KNN is $k=5$.

3.2 Data Set

Data of one building of the Universitat Politècnica de Catalunya (UPC) was obtained for this study: the bar of the ETSAV³ faculty in Sant Cugat of 150m². The energy consumptions of this building has been collected through a remote metering system every hour. Therefore there are 24 recordings per day.

The data set comprise a whole year of electricity consumption, from 13/11/2011 to 12/11/2012, with 91% for training and 9% for testing. The testing data comprises 35 different days (i.e. 35 test sets) distributed equally through the whole year; meaning around 9 days per season. And taking into account the seven days of the week (from Monday to Sunday). By choosing these days we pretend to evaluate the models against the changes caused by seasonal period(s) and day of the week.

The data used for testing the models is removed from the initial data set and it is replaced by missing values. Therefore, a training set of 8760 registers from which 864 (36 missing days x 24 hours per day) contain missing values. Depending on the configuration of the mask that has a maximum depth of 168 rows, the process to obtain the pattern rules can produce as much as 191 x 36 pattern rules containing one or more missing values. That is, 6876 patterns rules with missing values. Therefore, in the extreme case, from 7728 (7896-168) pattern rules, 6976 are discarded, and, therefore only 852 become available in the behaviour matrix to carry out the prediction.

3.3 Evaluation Criteria

There are many measures of forecast's accuracy in the literature [19]. In these experiments, we use the Mean Absolute Percentage Error (MAPE) to offer a forecasting performance from a multi-dimensional perspective. The reason for this choice is that MAPE can be used to compare the performance on different data sets, because it is a relative measure. However, it has to be highlighted that measures based on percentage errors have the disadvantage of being infinite or undefined if $Y_t = 0$ for any t in the period of interest. MAPE also has the disadvantage that puts a heavier penalty in negative errors than in positive errors [19]. MAPE is described in equation (6).

$$MAPE = 100 * \frac{1}{N} \sum_{t=1}^N |(y_r(t) - y_f(t)) / y_r(t)| \quad (6)$$

³ ETSAV: School of Architecture of the Vallès;

Table 3. MAPE Average (%) of the 35 test days representing an entire season year, in the Bar ETSAV in Sant Cugat in Barcelona, obtained by means of standard FIR Prediction and Flexible FIR Prediction (F1-2-3: output Forecast 1,2,3)

Sant Cugat Bar											
Depth: 24				Depth: 72				Depth: 48 (24+24)			
FIR	F1	F2	F3	FIR	F1	F2	F3	FIR	F1	F2	F3
Input Variables: 4											
123.61	120.48	120.26	120.28	94.20	92.60	92.25	93.02	95.96	94.21	94.14	93.57
Input Variables: 5											
80.56	79.36	79.34	79.34	46.63	46.24	45.97	46.12	65.93	66.28	65.23	66.34
Input Variables: 6											
15.28	15.44	15.39	15.50	13.95	14.81	14.90	15.25	16.01	15.85	15.83	16.48

3.4 Results

Table 3 shows the results obtained by each methodology for the three different depths studied and for the three experiments with different number of input variables. The lowest error in each experiment is highlighted. It is important to emphasize that the average error performed with the standard FIR does not contain those instances with the missing value problem and therefore cannot be predicted. The average prediction error of the standard FIR and our new implementation are equivalent and sometimes the error is even lower with Flexible FIR Prediction. These are promising results because although the new version of FIR tries to perform predictions with missing values, both in input pattern and behaviour matrix, the MAPE average is around 14% of error. As it was emphasize in Section 3, standard FIR does not contain those instances with the missing value, which means that results in standard FIR are not affected by wrong predictions with the missing value problem.

Table 4. MAPE Average (%) of the days where standard FIR Prediction could not predict due to missing values

Sant Cugat Bar									
Depth: 24			Depth: 72			Depth: 48 (24+24)			
F1	F2	F3	F1	F2	F3	F1	F2	F3	
Input Variables: 4									
111.76	109.27	109.41	28.03	21.95	35.43	46.72	45.53	35.47	
Input Variables: 5									
38.55	37.82	37.82	25.02	20.31	22.92	43.20	37.11	43.56	
Input Variables: 6									
15.95	14.97	16.86	14.14	14.56	16.32	14.72	14.58	17.83	

Table 5. Number of Instances (Inst) predicted by standard FIR Prediction and Flexible FIR Prediction

	Sant Cugat Bar					
	Depth: 24		Depth: 72		Depth: 48 (24+24)	
	FIR	Flex FIR	FIR	Flex FIR	FIR	Flex FIR
Input Variables: 4						
Inst	809	840	814	840	813	840
%	96.31	100.00	96.90	100.00	96.79	100.00
Input Variables: 5						
Inst	816	840	814	840	756	840
%	97.14	100.00	96.90	100.00	90.00	100.00
Input Variables: 6						
Inst	795	840	691	840	708	840
%	94.64	100.00	82.26	100.00	84.29	100.00

The main reason why high error values appear in some cells of Table 3 is because the table shows the average of 35 test days. If in those cases the prediction of a single day is quite bad then the average error is considerably increased.

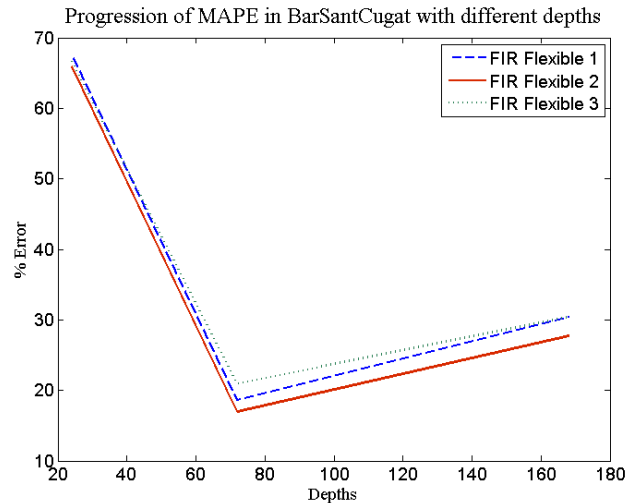


Figure 7. Error progression with depths 24, 72 and 162(24+24) in the three output forecast models

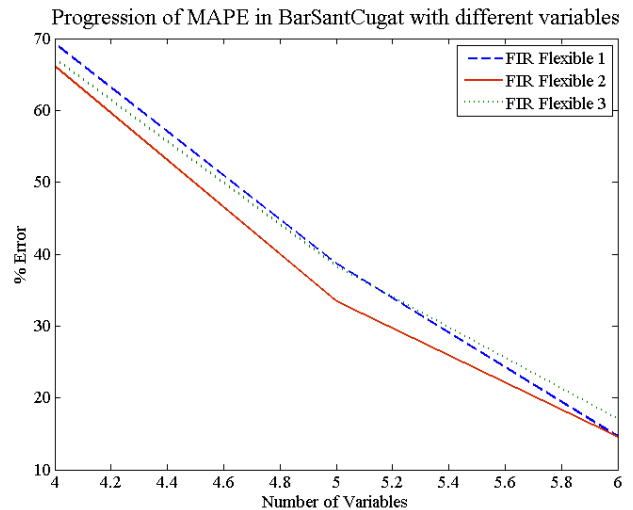


Figure 8. Error progression with input variables 4, 5 and 6 in the three output forecast models

Table 4 shows the results obtained only when standard FIR was not able to perform a prediction, i.e. when there were missing values in the input pattern and/or the input pattern was not found in the behaviour matrix. When Table 4 is analysed, it can be seen that the MAPE average for those instances (registers) that contain missing values is around 14% in the best cases, which is a similar to the average error prediction of Table 3, although missing information is present.

In Table 5 the total instances predicted in the experiments by the different methodologies are shown. When the Flexible FIR Prediction was used, all the instances (from the test dataset) were predicted, while for the standard FIR decreases up to 82.26% in some cases.

Both Figure 6 and Figure 7, represent the MAPE average against the depth of the mask and number of input variables respectively, with different forecast computations. FIR Flexible 1 corresponds to F1, FIR Flexible 2 to F2 and so on.

4. CONCLUSIONS AND FUTURE WORK

In this paper an improved version of the Fuzzy Inductive Reasoning (FIR) is proposed. It is able to cope with the problem of missing values in the input pattern, as well as learn from instances with missing values in the behaviour matrix.

The experiments presented contain missing values during the training process and the online prediction. This new technique is able to cope with them and take advantage of the inherent information in the data, even though it contains missing values.

In addition, in almost real-time applications the fact that a prediction cannot be performed is a weakness because it may crash the whole system. With Flexible FIR Prediction we reduce or finish with this issue.

Based on Figure 6 and Figure 7 it is clear that, with this data, the optimal configuration of Flexible FIR is when mask depth is 72 and the number of variables is 6. Furthermore, FIR Flexible 2 (F2), in general, predicts better than the other two proposed output forecast models.

As future work, we plan to extend our experiments to several public buildings and residential houses to demonstrate the scalability of our technique. It is also planned to define more output forecast models that consider, for instance, the causal relevance of the relaxed parameters in the Euclidian distance, or assign confidence levels to the predictions performed with relaxed input patterns and behaviour matrices. Finally, another future work is to integrate this improved version of FIR in a next generation of smart home energy management solution, called Local Decision Support System (LDSS) and deployed in two European cities: Plovdiv and Rijeka.

5. ACKNOWLEDGMENTS

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement number 608712, project iURBAN.

6. REFERENCES

[1] T. E. E. The commission to the European Parliament, the Council, S. committee, and the committee of the regions. Energy roadmap 2050 [com/2011/885]. Technical report.

[2] European Commission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity [com/2014/0356 final]. Technical report.

[3] "With New Factory, Tesla Ventures Into Solar Power Storage for Home and Business." *The New York Times* (May 1, 2015), sec. B p.3.

[4] Douglas, A. P., Breipohl, A. M., Lee, F. N. and Adapa, R. 1988. Risk due to load forecast uncertainty in short term power system planning. *IEEE Trans. Power Syst.*, vol. 13(4), pp. 1493–1499.

[5] Hobbs, B. F., Jitprapaikularn, S., Konda, S., Chankong, V., Loparo, K. A. and Maratukulam, D. J. 1999. Analysis of the value for unit commitment of improved load forecasting. *IEEE Trans. Power Syst.*, vol. 14(4), pp. 1342–1348. □

[6] Huina, M., Xiao-Jun, Z., Gang, L., Yong-Jie, Z. and Keane, J. A. 2009. Short-term and midterm load forecasting using a bilevel optimization model. *IEEE Trans. Power Syst.*, vol. 24(2), pp. 1080–1090.

[7] Jones, R.H. 1980. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, vol. 22, pp. 389–395.

[8] Kohn, R., Ansley, C.F. 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*. vol. 81, pp. 751–761.

[9] Gómez, V., Maravall, A., Peña, D. 1999. Missing observations in ARIMA models: skipping approach versus additive outlier approach. *Journal of Econometrics*. vol. 88, pp. 341–363.

[10] Cipra, T., Romera, R. 1997. Kalman filter with outliers and missing observations. *Test* 6. pp. 379–395.

[11] Kohn, R., Ansley, C.F., 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*. vol. 81, pp. 751–761.

[12] Bermúdez, J. D., Corberán-Vallet, A., Vercher, E. 2009. Forecasting time series with missing data using Holt's model. *Journal of Statistical Planning and Inference*. vol. 139, pp. 2791–2799.

[13] Nebot, A. and Mugica, F. 2012. Fuzzy Inductive Reasoning: a consolidated approach to data-driven construction of complex dynamical systems. *International Journal of General Systems*, vol. 41(7), pp. 645–665.

[14] Hinojosa, V. H., Hoese A. 2010. Short-Term Load Forecasting Using Fuzzy Inductive Reasoning and Evolutionary Algorithms. *IEEE Transactions on power systems*. vol. 25 (1).

[15] Klir, J., Elias, D. (2002). *Architecture of Systems Problem Solving*, 2nd. Ed., Plenum Press, New York.

[16] Nebot, A., Mugica, F., Cellier, F., Vallverdú, M. 2003. Modeling and Simulation of the Central Nervous System Control with Generic Fuzzy Models, *Transactions of The Society for Modeling and Simulation*, vol. 79(11), pp. 648–669.

[17] Escobet, A., Nebot, A., Cellier, F.E. 2008. Visual-FIR: A tool for model identification and prediction of dynamical complex systems, *Simulation Modelling Practice and Theory*, vol. 16, pp. 76–92.

[18] Law, A.M. and Kelton, W.D. 1991. *Simulation Modeling and Analysis*, vol. 2, McGraw-Hill.

[19] Hyndman, R.J., Koehler, A.B. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, pp. 679–688.

[20] Jurado, S., Nebot, A., Mugica F. 2015. Hybrid methodologies for electricity load forecasting: Entropy-Based Feature Selection with Machine Learning and Soft Computing Techniques. *Energy*. (accepted but pending to be published)

