FORECASTING KNOWLEDGE EXTRACTION BY
COMPUTATIONAL INTELLIGENCE TECHNIQUES

BY

ELIA GEORGIANA DRAGOMIR¹ and MIHAELA OREA²

Petroleum - Gas University of Ploiești,
¹Faculty of Letters and Science
²Faculty of Mechanical and Electrical Engineering

Received: September 10, 2014
Accepted for publication: October 17, 2014

Abstract. Most of the recently developed intelligent systems have a
c knowledge base that incorporates the expertise domain knowledge and use it in
reasoning chains during decision making. An important problem that should be
solved by an intelligent system is forecasting the evolution of specific parameters
that are monitored, for example. Among the various approaches that provide an
efficient solution to the forecasting problems, some computational intelligence
techniques allow the extraction of the forecasting knowledge under the IF-THEN
rules form. The paper presents a general methodology that can be used to
forecasting knowledge extraction and the experimental results of a comparative
study between a computational intelligence technique, the adaptive neuro-fuzzy
inference system (ANFIS), and a decision tree based technique, CART, applied to
air pollution forecasting rules extraction, by following the proposed methodology.

Key words: computational intelligence; forecasting knowledge extraction;
adaptive neuro-fuzzy inference system; methodology.

2010 Mathematics Subject Classification: 68T27, 68T30.

¹Corresponding author; e-mail: elia.dragomir@yahoo.com
1. Introduction

The development of an intelligent system requires knowledge extraction from different sources related to the domain of expertise. An important problem that should be solved by any intelligent system is forecasting the evolution of specific parameters that are monitored, for example.

Nowadays, a various range of models can be used in forecasting knowledge extraction based on historical data (past measurements) from the traditional methods that include statistical analysis to the computational intelligence techniques which can search in large volumes of data and find hidden patterns that can be used in a decision making process.

The computational intelligence methods are data oriented and computationally efficient. Thus, large amount of data can be processed in shorter time and can be integrated into existing information systems for different tasks such as forecasting knowledge extraction.

The paper presents a general methodology that can be used to forecasting knowledge extraction and discusses the experimental results of a comparative study between an adaptive neuro-fuzzy inference system (ANFIS) and a decision tree based technique, CART, applied to air pollution forecasting rules extraction, by following the proposed methodology.

The paper is organized as follows. Section 2 makes a brief presentation of the computational intelligence techniques. The general forecasting knowledge extraction methodology is introduced in section 3. The results of a comparative study performed between the ANFIS model and a decision tree technique are discussed in section 4. The last section concludes the paper.

2. Computational Intelligence Techniques

The term of computational intelligence (CI) is known as a collection of intelligent computational methodologies that solve complex problems that traditional methods like statistical models cannot (Palit & Popovic, 2005). Different computational models and tools are included in CI: granular computing, neural computing, and evolutionary computing. They can be applied with success to uncertain and imprecise data sets.

The main computational intelligence techniques are artificial neural networks (ANN), fuzzy systems, evolutionary algorithms (e.g. genetic algorithms, genetic programming), and swarm intelligence (e.g. ant colony optimization - ACO, particle swarm optimization - PSO). Also, combinations of these techniques have been proposed in the literature. A successful one is the adaptive neuro-fuzzy inference system (ANFIS) that has the ability to combine the expert’s empirical knowledge, in the form of IF-THEN rules, through the data by using a learning algorithm (Jang, 1993).
3. Forecasting Knowledge Extraction Methodology

We have synthesized the main steps that should be followed during forecasting knowledge extraction under a general methodology described as follows (Dragomir, 2014).

**General Methodology (Forecasting Knowledge Extraction)**

**Input:** historical databases  
**Output:** the forecasted knowledge (Knowledge Base - KB)  
**Steps:**

1. select the significant input parameters \((x_i)\) for the forecasted variable \((y)\) by using a certain selection method (e.g. Principle Components Analysis – PCA);
2. create the current databases from the given historical databases, by using the result of the previous step;
3. databases pre-processing (missing values filling, data cleaning, data transformation, data integration, data reduction);
4. apply the forecasting knowledge extraction algorithm (e.g. ANFIS, ANN, or a decision tree based one such as ID3, C4.5, CART) and provide the forecasting knowledge;
5. add to or create a knowledge base (KB) with the forecasting knowledge derived in step 4.

Fig. 1 shows the main steps of the proposed methodology.

![Fig. 1 – The main steps of the methodology.](image-url)
We propose the use of the Principal Component Analysis in order to reduce the input vector dimensionality. This technique can determine the principal components that contribute to the generated output and, therefore, can select the most important input parameters for the forecasting knowledge extraction method.

Step 3 is important because low quality data set generate low quality forecasting rules. Various pre-processing techniques can be applied to the databases: missing values filling, data cleaning, data transformation, data integration, data reduction (Han & Kamber, 2006). The filling of the missing values can be done by ignoring the tuple, fill in the missing value manually, using a global constant, the attribute mean or the most probable value etc. Data transformation consists in one of the methods: normalization, smoothing, aggregation, generalization of the data and strategies for data reduction include: data compression, data cube aggregation or dimension reduction.

Fig. 2 presents the forecasting knowledge extraction as a block diagram.

Forecasting knowledge extraction approaches are computational intelligence techniques and data mining. Over the last decades, various data mining methods have been developed to extract forecasting knowledge from huge databases, such as fuzzy logic based computing, neurocomputing and evolutionary computing. Among these, fuzzy inference systems (e.g. ANFIS), artificial neural networks, decision trees (e.g. CART) can perform forecasting knowledge extraction in the form of production rules due to the possibility of dealing with imprecise and uncertain data along with a large number of dependent and independent interactions among variables which are often recorded in real time systems, resulting in highly complex non-linear dynamics. Furthermore, there is a practical usage of forecasting knowledge extraction techniques in various domains like: environmental and business systems, engineering and physical science.
3.1. Brief State of the Art

A review of the recently reported research work in the area of forecasting knowledge extraction revealed a lack of a general methodology that should be followed. In this sense, our methodology provides the main guidelines for any domain of application.

The most successfully used forecasting knowledge extraction methods are ANFIS and CART and our brief state of the art is concentrated on them. Over the last decade, ANFIS and CART were applied by different researchers, being appropriate to construct hydrological time-series forecasting systems (Nayak et al., 2005; Firat & Güngör, 2008; Valenca & Ludermir, 2000), to design air pollution and meteorological knowledge extraction and forecasting systems (Oanh et al., 2010; Dutot et al., 2007). There are studies in literature which applied ANFIS and CART for load forecasting (Ding, 2006; Hanmandlu & Chauhan, 2011), power supply chain predictions (Mellit & Kalogirou, 2011; Melin et al., 2012). Many experiments conclude that these techniques can be successfully used by the enterprisers to improve their activities: product development knowledge extraction (Arafeh et al., 1999; Vairappan et al., 2009), sales (Kuo & Xue, 1999; Wang & Chen, 2008).

3.2. Case Study

We have performed a comparative study between a CI technique, ANFIS, and a decision tree based technique, CART, applied to air pollution forecasting knowledge extraction, by following the proposed methodology. We have chosen as a case study an environmental problem, air pollution forecasting. Air pollution data are difficult to model mainly due to the non-linearity of the chemical atmospheric processes. Another characteristic of this domain is the lack of sufficient data or bad quality data. Therefore, the forecasting knowledge extraction is difficult to be model with traditional techniques.

3.3. The Forecasting Problem

The problem goal is to forecast the ozone (O3) concentration based on the historical recorded data (concentrations) for other important air pollutants: carbon monoxide (CO), sulphur dioxide (SO2), nitrogen dioxide (NO2), nitrogen monoxide (NO). For the ANFIS and CART performance evaluation we have used two statistical parameters: the root mean square error (RMSE) and the correlation coefficient (R2).

3.4. The Data Set

For this study we have used the hourly recorded data for the air pollutants concentrations at a monitoring station from the National Air Quality
Elia Georgiana Dragomir and Mihaela Oprea

Monitoring Network located in the East side of the Ploiești city. The data set includes data recorded for a period of 12 months, from January 2011 to December 2011. We have split the dataset in three datasets for: training, testing, and validation. Based on the air pollution indices used at the national level, we have divided the numerical recorded data for each air pollutant into six major intervals (excellent, veryGood, good, moderate, poor, severe), where excellent label reflects that the parameters values are low and severe label indicates an important pollution episode.

3.5. The ANFIS Model

3.5.1. The ANFIS Architecture

In the first step of the proposed methodology the most appropriate input parameters were selected. In order to determine the most influential input attribute in predicting the output there were testing the possibilities of having as inputs 1, 2, 3 or 4 nodes. The results show that the most appropriate combination is between two parameters carbon monoxide and nitrogen monoxide (RMSE error is 0.05), presented in Fig. 4 (left). Therefore, in this case study we have used these parameters as inputs and ozone as output.

In the second step there were extracted only the carbon monoxide, nitrogen monoxide and ozone recorded data from the initial database. This step was followed by the pre-processing phase. All the missing data were replaced by the nearby values average.

To apply the ANFIS forecasting knowledge extraction algorithm a hybrid method consisting of backpropagation for the input parameters and least square estimation for the parameters associated with the output membership functions were applied. The proposed architecture of ANFIS model for the ozone prediction is presented in Fig. 3.

The ANFIS detailed architecture for two input variables, one carbon monoxide with fuzzy set A1={excellent}, A2={veryGood}, A3={good}, A4={moderate}, A5={poor} and A6={severe} and the other one nitrogen monoxide with fuzzy set B1={excellent}, B2={veryGood}, B3={good}, B4={moderate}, B5={poor} and B6={severe} is indicated in Fig. 4 (right).

Fig. 3 – The ANFIS architecture.
Fig. 4 – The training and testing errors for the two inputs possibilities (left) The ANFIS detailed structure (right).

It has 5 layers as follows: the parameters of the first layer are referred to as premise that gives the degree of Fuzzy membership of the input. The second layer contains the nodes which represents the firing strength of a fuzzy rule as the output nodes of the third layer are called normalized firing strengths. On the fourth layer every node is a linear combination of the input variables and the single node of the fifth layer computes the output as the sum of all incoming values (Jang, 1993).

The computations of the membership function parameters are facilitated by a gradient vector which provides a measure of how well the FIS system is modelling the input/output data (Fuzzy logic toolbox user's guide for use with MATLAB, 2010). The numbers of linear parameters and non-linear parameters were found to be 108 and 36 respectively.

3.5.2. Rule Extraction

ANFIS creates 36 rules, as shown in Fig. 5 (left) during the training phase. Each rule has a certain weight that can have any value between 0 and 1. 0 means that in the final output that rule has no importance, and 1 indicates that rule is very strong for the final output (Zadeh, 1989).

In Table 1 there are presented some of the IF-THEN rule extracted by the ANFIS model. The IF clauses are linked using the logical operator AND and cover all the possible combinations among the input data.

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Rule Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IF (CO is excellent) and (NO is excellent) THEN (O3 is excellent)</td>
</tr>
<tr>
<td>9</td>
<td>IF (CO is very good) and (NO is good) THEN (O3 is good)</td>
</tr>
<tr>
<td>14</td>
<td>IF (CO is good) and (NO is bad) THEN (O3 is bad)</td>
</tr>
<tr>
<td>22</td>
<td>IF (CO is severe) and (NO is good) THEN (O3 is severe)</td>
</tr>
</tbody>
</table>
In Fig. 5 (right) there can be seen entire fuzzy inference process, right from how the membership functions are being satisfied in every rule to how the final output is being generated through defuzzification (Fuzzy logic toolbox user's guide for use with MATLAB, 2010). It can be simulated the model response for specific input values.

![Fig. 5 – The rules generated by the ANFIS model (left)
The Rule viewer for ANFIS model (right).](image)

3.6. The Decision Tree Based Model

3.6.1. The CART Architecture

Decision trees are very popular methods for classification and prediction, being easily interpreted compared with other machine learning techniques. The main components of a decision tree enrol decision nodes and leaf nodes. A decision tree allows the extraction of the forecasting knowledge under the IF-THEN rules form. CART algorithm (Classification and Regression Trees, (Han & Kamber, 2006), designed by Breiman in 1984, is a method that creates a binary tree using the Twoing selection criteria and a pruning method based on the cost complexity.

In this case study, the training and testing data sets for the CART model were used similar to the ANFIS model.

The Principal Component Analysis is used in the first step of the proposed methodology in order to find the relevant parameters for ozone concentration forecasting. The input vector is first normalized and afterwards the Principal Components Analysis eliminates those components that have the least contribution to the variation in the data set (Fuzzy logic toolbox user's guide for use with MATLAB, 2010). As a result there were used only two principal components as inputs in the decision tree model (carbon monoxide and sulphur dioxide).
The second and the third steps of the methodology are applied similar to the ANFIS model. In the forth step the CART algorithm is used to generate the decision tree as presented in Fig. 6.

The number of nodes created for the decision tree is 3 and the number of leafs is 4.

Fig. 6 − The CART decision tree architecture.

3.6.2. Rule Extraction

Reading the decision tree from root toward the leaf nodes the IF – THEN rules are created which allow the extraction of the forecasting knowledge. There can be observed that applied to this database the CART model does not cover all the possible values for ozone concentration (the excellent and good ozone concentration values cannot be determined using CART algorithm).

<table>
<thead>
<tr>
<th>Rule number</th>
<th>Rule Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IF (CO is excellent or veryGood) and (SO2 is excellent or veryGood or good) THEN (ozone is veryGood)</td>
</tr>
<tr>
<td>2</td>
<td>IF (CO is excellent or veryGood) and (SO2 is moderate or poor or severe) THEN (ozone is poor)</td>
</tr>
<tr>
<td>3</td>
<td>IF (CO is not excellent or veryGood) and (SO2 is excellent or veryGood or good) THEN (ozone is moderate)</td>
</tr>
<tr>
<td>4</td>
<td>IF (CO is not excellent or veryGood) and (SO2 is good or moderate or poor or severe) THEN (ozone is severe)</td>
</tr>
</tbody>
</table>

4. Results and Discussions

In Table 3 there are centralized information about both ANFIS and CART models built in order to extract forecasting knowledge under the form of IF-THEN rules. Even though both models have 2 inputs, these are different
from one model to another: for the ANFIS the carbon monoxide and nitrogen monoxide are selected as most influential parameters and for the CART model, carbon monoxide and sulphur dioxide are used.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>The Principal Parameters of the ANFIS and CART Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANFIS</td>
</tr>
<tr>
<td>Number of inputs</td>
<td>2</td>
</tr>
<tr>
<td>Number of nodes</td>
<td>101</td>
</tr>
<tr>
<td>Number of rules</td>
<td>36</td>
</tr>
<tr>
<td>Time to build model</td>
<td>0.34 s</td>
</tr>
</tbody>
</table>

The number of nodes and rules are significantly much higher for the ANFIS model but the precision is more accurate. There are ANFIS rules that can cover all possible situations comparing with the CART response. The time needed by each model to be built is lower for ANFIS model against CART.

The performance of those models in terms of RMSE and \( R^2 \) are shown in Table 4. The RMSE values for ANFIS model are 0.03 in the training phase and 0.05 in the testing step. These values are lower than CART’s values. Comparing the \( R^2 \) parameter for both models there can be seen better results for the ANFIS: 0.96 for training data set and 0.91 for testing step. Therefore the ANFIS model performs better in this experiment.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Performance of ANFIS and CART for Training and Validation Phases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>RMSE Training Testing</td>
</tr>
<tr>
<td>ANFIS</td>
<td>0.03 0.05</td>
</tr>
<tr>
<td>CART</td>
<td>0.15 0.17</td>
</tr>
</tbody>
</table>

In Fig. 7 (left) it is presented the ANFIS error curve for 500 epochs: the green line represents the training data set and the red one the testing data. The black circle indicates the minimal error value which occurs around the 300\(^{th}\) epoch. After 350\(^{th}\) epoch the testing data line goes up which means that further training overfits the data. Fig. 7 (right) sketches the inputs – output snapshot of the ANFIS model at the minimal testing error during the training process. The upper-right corner reflects the direct influence of nitrogen monoxide over the ozone concentration.
In the case of forecasting the ozone concentration in Ploiesti the ANFIS model is more appropriate than CART model having a better prediction accuracy and better generalization capability in comparison to the CART model.

5. Conclusions

Computational intelligence provides powerful forecasting knowledge extraction techniques. We have made a comparative analysis between a computational intelligence technique, ANFIS, and a decision tree based technique, CART, applied to air pollution forecasting rules extraction. The experimental results revealed that the ANFIS model performs better than the decision tree based technique.

REFERENCES

* Fuzzy logic toolbox user's guide for use with MATLAB, 2010.


**EXTRAGEREĂ CUNOȘTINȚELOR DE PREDICȚIE UTILIZÂND TEHNICI DE INTELIGENȚĂ COMPUTAȚIONALĂ**

(Rezumat)

Multe dintre sistemele inteligente dezvoltate recent conțin o bază de cunoștințe ce încorporează cunoștințele experților din domeniu și le utilizează prin mecanisme de inferență în procesul de decizie. O problemă importantă este de a rezolva o predicție de către un sistem inteligent este predicția evoluției unor anumite parametri monitorizați. Prin abordări numeroase de abordări de pot da o soluție eficientă în problemele de predicție, unele tehnici de calculația permit extragerea cunoștințelor de predicție sub forma unor reguli de forma DACĂ-ATUNCI. Acest articol prezintă o metodologie generală care poate fi utilizată în extragerea cunoștințelor de predicție și rezultatele experimentale ale unui studiu comparativ între o tehnică de inteligență computațională, sistem de inferență adaptiv neuro-fuzzy (ANFIS) și o tehnică bazată pe arbor de decizie, CART, aplicată în predicția poluării aerului, urmărind metodologia propusă.