A Constructive-Fuzzy System Modeling for Time Series Forecasting

I. Luna¹, R. Ballini² and S. Soares¹

¹Department of Systems, School of Electrical and Computer Engineering, University of Campinas - Brazil
²Institute of Economy, University of Campinas - Brazil
Summary

• Introduction;
• Time series analysis: data pre-processing;
• Input selection;
• General structure of a fuzzy rule-based model;
• Constructive learning;
• Case study: NN3 competition;
• Conclusions and future works.
Time series modeling

Figure 1: Time series modeling.
Analysis and pre-processing

• Reduced data set of the NN3 competition;
• Stationarity: required for input selection;
• Seasonal and trend components:


• Trend component: series 1, 5 and 9;
• All series with no trend were transformed according to:

\[ z^k(m) = \frac{y^k(m) - \mu(m)}{\sigma(m)} \]

\( z^k \): stationary version of the time series;
\( y^k, \; k = 1, \ldots, k \)-th observation;
\( \mu(m) \) is the monthly average value and \( \sigma(m) \) is the monthly standard deviation.
Input selection

1. **FNN** (*False Nearest Neighbors*): determines the minimum number of lags necessary to represent each pattern or *state* of the time series;

2. **PMI** (*Partial Mutual Information*): measure of information that each new variable $x$ provides, taking into account an existing set of inputs $Z$. Given variables $X$ e $Y$, PMI score between $X$ and $Y$ is defined by:

$$PMI = \frac{1}{N} \sum_{i=1}^{N} \log_e \left[ \frac{f_{X',Y'}(x_i', y_i')}{f_{X'}(x_i') f_{Y'}(y_i')} \right]$$

where:

$$x_i' = x_i - E(x_i|Z) \quad \text{e} \quad y_i' = y_i - E(y_i|Z)$$

$Z$ is the set of inputs already chosen. $E(\cdot|Z)$ is the conditional expected value; $N$ is the number of input-output patterns.
Input selection

\[ y = x^4 + e_1 \]
\[ x = \text{sen}(2\pi t/T) + e_2 \]

\( T = 20; \ t = 1, \ldots, 200; \)  
\( e_1 \) and \( e_2 \) are noisy signals with normal distribution, \( \mu = 0 \) and \( \sigma = 0, 1. \)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI</td>
<td>0.4199</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.0032</td>
</tr>
</tbody>
</table>
Input selection

Number inputs  Complexity

Local minimum

Figure 2: FNN.

Figure 3: PMI.
A general structure

Figure 4: A general structure of a fuzzy-rule based system, composed by a total of $M$ fuzzy rules.
A general structure

- \( \mathbf{x}^k = [x_1^k, x_2^k, \ldots, x_p^k] \in \mathbb{R}^p \) is the input vector at instant \( k, k \in \mathbb{Z}_0^+ \);
- \( \hat{y}^k \in \mathbb{R} \) is the estimate output;
- Given centers \( \mathbf{c}_i \in \mathbb{R}^p \) and covariance matrices \( \mathbf{V}_i \ i = 1, \ldots, M \), membership degrees \( g_i(\mathbf{x}^k) \) are defined as:

\[
 g_i(\mathbf{x}^k) = g_i^k = \frac{\alpha_i \cdot P[ i | \mathbf{x}^k ]}{\sum_{q=1}^{M} \alpha_q \cdot P[ q | \mathbf{x}^k ]} 
\]

(2)

with \( \alpha_i \geq 0, \sum_{i=1}^{M} \alpha_i = 1 \) and:

\[
 P[ i | \mathbf{x}^k ] = \frac{1}{(2\pi)^{p/2} \det(\mathbf{V}_i)^{1/2}} \times
\]

\[
 \times \exp \left\{ -\frac{1}{2} (\mathbf{x}^k - \mathbf{c}_i) \mathbf{V}_i^{-1} (\mathbf{x}^k - \mathbf{c}_i)^T \right\} 
\]

(3)
A general structure

- Each local model $y_i^k, i = 1, \ldots, M$ is estimated by a linear one:

$$y_i^k = \phi^k \times \theta_i^T$$

where $\phi^k = [1 \ x_1^k \ x_2^k \ \ldots \ x_p^k]$ and $\theta_i = [\theta_{i0} \ \theta_{i1} \ \ldots \ \theta_{ip}]$.

- The output model $\hat{y}^k$ is computed as:

$$\hat{y}^k = \sum_{i=1}^{M} g_i(x_i^k) y_i^k$$

(5)
Constructive learning

Initialization

- **E step:** $g_i^k$ is estimated given $x^k$ and $y^k \Rightarrow posterior$ estimate $h_i^k$;

$$h_i^k = \frac{\alpha_i P(i \mid x^k)P(y^k \mid x^k, \theta_i)}{\sum_{q=1}^{M} \alpha_q P(q \mid x^k)P(y^k \mid x^k, \theta_q)}$$

Adaptation

- **M step:**
  - Model parameters are adjusted;
  - Adding and pruning conditions are verified.

$$\alpha_i = \frac{1}{N} \sum_{k=1}^{N} h_i^k$$ (6)
Second phase: Adaptation

- **Adding a new rule**: Assuming a normal input data distribution, with a confidence level equal to $\gamma\%$:

  \[ \Omega \text{ is an i/o data set so that } \max_{i=1,\ldots,M} (P(i|\mathbf{x}^k \in \Omega)) < 0.14 \]

  \[ \Rightarrow \text{If } N_\Omega > 0 \]

  Then create a new rule.

  \[ \gamma \approx 73\% \]

  \[ z_\gamma \approx 1.1 \]

  \[ \approx 0.14 \approx (1 - 0.73)/2 \]

- **Pruning a new rule**: $\alpha_i$ is proportional to the sum of all $h_{ik}$. Thus, the more times the rule is strongly activated, the higher its $\alpha_i$ will be.

  \[ \Rightarrow \text{If } \alpha_i < \alpha_{\text{min}}, \text{ then the } i\text{-th rule will be pruned.} \]
Case study: NN3 competition

- Reduced data set;

Table 1: Global prediction errors for series NN3_102 and NN3_104.

<table>
<thead>
<tr>
<th>Series</th>
<th>In sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 step ahead</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>NN3_102</td>
<td>$k = 4, \ldots, 108$</td>
</tr>
<tr>
<td>NN3_104</td>
<td>$k = 4, \ldots, 97$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series</th>
<th>Out of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 step ahead</td>
</tr>
<tr>
<td></td>
<td>1 to 18 steps ahead</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>NN3_102</td>
<td>$k = 109, \ldots, 126$</td>
</tr>
<tr>
<td>NN3_104</td>
<td>$k = 98, \ldots, 115$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Series</th>
<th>$M$</th>
<th>sMAPE (%)</th>
<th>MAE ($u$)</th>
<th>sMAPE (%)</th>
<th>MAE ($u$)</th>
<th>sMAPE (%)</th>
<th>MAE ($u$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN3_102</td>
<td>3</td>
<td>3.41</td>
<td>179.36</td>
<td>4.72</td>
<td>287.98</td>
<td>11.40</td>
<td>658.05</td>
</tr>
<tr>
<td>NN3_104</td>
<td>3</td>
<td>10.98</td>
<td>438.27</td>
<td>6.75</td>
<td>334.93</td>
<td>12.32</td>
<td>612.32</td>
</tr>
</tbody>
</table>
Case study: NN3 competition

Figure 5: Multiple steps ahead: (a) autocorrelation coefficients estimates, (b) predictions for series NN3_102.
Figure 6: Multiple steps ahead: (a) autocorrelation coefficients estimates, (b) predictions for series NN3_104.
Case study: NN3 competition

Table 2: Some characteristics of input selection and model construction.

<table>
<thead>
<tr>
<th>Time series</th>
<th>Difference</th>
<th>Num. inputs</th>
<th>Inputs (lags)</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1, 2, 4</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1, 3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1, 10</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>3</td>
<td>1, 2, 3</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1, 2</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>2, 3, 4</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>2</td>
<td>2, 3</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>2</td>
<td>1, 6</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Case study: NN3 competition

Figure 7: One and multi-step ahead forecasting for time series NN3_101 to NN3_104.
Case study: NN3 competition

Figure 8: One and multi-step ahead forecasting for time series NN3_105 to NN3_108.
Case study: NN3 competition

Figure 9: One and multi-step ahead forecasting for time series NN3_109 to NN3_111.
Conclusions and Future works

• This work presents a methodology for time series modeling.
• Statistical tools combined with novel methodologies provide adequate models.

• Objectives achieved:
  • The study of the different tasks that compose the methodology: from data pre-processing to model validation.
  • The automatic selection of a suitable model structure;

• What needs to be improved:
  • Initialization phase;
  • Adding and pruning conditions.

Thanks for your attention.

Ivette Luna

iluna@cose.fee.unicamp.br