Identification of fuzzy regression models

A. Bisserier, S. Galichet, R. Boukezzoula

LISTIC

13 Novembre 2008
1. Introduction
2. Notations and concepts
3. Fuzzy linear regression
4. Proposed approach
5. Illustrative examples
6. Conclusion
Introduction

Crisp data

Conventional optimization techniques

Conventional regression:
least squares

Model identification
Relation between inputs and outputs

Observed data
nature

Crisp data

Conventional optimization techniques

Conventional regression:
least squares
...
Introduction

Model identification
Relation between inputs and outputs

Observed data nature

Crisp data
Conventional optimization techniques
Conventional regression: least squares...

Imprecise and/or uncertain data
Fuzzy optimization techniques
Fuzzy regression: fuzzy least squares...

A. Bisserier (LISTIC)
Identification of fuzzy regression models
13 Novembre 2008
Introduction

Model identification
Relation between inputs and outputs

Observed data nature

Crisp data
Conventional optimization techniques
Conventional regression: least squares ...

Imprecise and/or uncertain data
Fuzzy optimization techniques
Fuzzy regression: fuzzy least squares ...

General problem
- What kind of fuzzy models have to be used?
- What kind of data have to be manipulated for the identification process?
Introduction

Model identification
Relation between inputs and outputs

Observed data nature

Crisp data
Conventional optimization techniques
Conventional regression: least squares
...

Imprecise and/or uncertain data
Fuzzy optimization techniques
Fuzzy regression: fuzzy least squares
...

General problem
- What kind of fuzzy models have to be used?
- What kind of data have to be manipulated for the identification process?
⇒ How can we identify a fuzzy model on this kind of data?
Considered model: multi-inputs, single-output

- Objective: to determine a relationship $Y = h(x)$ between inputs and output
- $h$ considered as linear $\rightarrow$ model of the form:

$$Y = \sum_{i=1}^{N} A_i \cdot x_i$$

- Fuzzy model: fuzzy coefficients $A_i$
Notations and concepts

Observed data for the identification

Set of $M$ observed data:
The $j^{th}$ data:
- a crisp inputs vector $x_j = (x_{0j}, x_{1j}, ..., x_{Nj})$
- the corresponding fuzzy output $Y_j$
  $\rightarrow$ a symmetrical triangular fuzzy number

Remark: choice made for the sake of simplicity
Conventional intervals

- Set of elements in $\mathbb{R}$ between a lower and an upper bound

$$a = \{x \mid a^- \leq x \leq a^+, x \in \mathbb{R}\}$$

- Midpoint and Radius

$$M(a) = M_a = (a^- + a^+)/2$$

$$R(a) = R_a = (a^+ - a^-)/2$$

calculus on sets
Notations and concepts

Conventional intervals

- Set of elements in $\mathbb{R}$ between a lower and an upper bound

\[ a = \{ x | a^- \leq x \leq a^+, x \in \mathbb{R} \} \]

- Midpoint and Radius

\[ M(a) = M_a = (a^- + a^+)/2 \]
\[ R(a) = R_a = (a^+ - a^-)/2 \]

- calculus on sets

Conventional intervals: inclusion

- general relationship:

\[ a \subseteq b \Leftrightarrow |M(b) - M(a)| \leq R(b) - R(a) \]
Notations and concepts

**Fuzzy intervals**

- Horizontal AND vertical dimension

⇒ intervals at two levels:
  - kernel: \( K_A = [K_A^-, K_A^+] \)
  - support: \( S_A = [S_A^-, S_A^+] \)

- Kernel included in the Support

- function linking the two levels: profiles

- \( \alpha \)-cut ⇒ conventional interval
Notations and concepts

Fuzzy intervals: linear profiles

- **general case**: trapezoidal fuzzy interval
  \[
  A = (K_A, S_A) = ([K_A^-, K_A^+], [S_A^-, S_A^+])
  \]

- **a particular case**: symmetrical triangular fuzzy interval
  \[
  A = (K_A, R_A)
  \]
Notations and concepts

- inclusion of two fuzzy intervals: general definition
  \[ A \subseteq B \iff \forall x, \mu_A(x) \leq \mu_B(x) \]

- in the case of trapezoidal fuzzy intervals:
  \[ A \subseteq B \iff K_A \subseteq K_B \text{ et } S_A \subseteq S_B \]

- in the case of symmetrical triangular fuzzy intervals:
  \[ \Rightarrow \text{ equality of the Kernel values} \]
Fuzzy linear regression

Two distinct approaches

Diamond

- Minimization of the model output quadratic error
- Minimization of a distance between fuzzy numbers
- Search for the model most appropriated with data

Tanaka

- Minimization of the model output uncertainty
- Optimization of a criterion under constraints
- Search for the less uncertain model respecting the constraints
Fuzzy linear regression

Minimization of the model output quadratic error
Minimization of a distance between fuzzy numbers
Search for the model most appropriated with data

Which constraints?

Relationship between observed and predicted outputs:

- **necessity** model: predicted outputs included in observed ones
- **conjunction** model: no empty intersection between predicted and observed outputs

Two distinct approaches

Diamond

Minimization of the model output quadratic error
Minimization of a distance between fuzzy numbers
Search for the model most appropriated with data

Tanaka

Minimization of the model output uncertainty
Optimization of a criterion under constraints
Search for the less uncertain model respecting the constraints
Fuzzy linear regression

Two distinct approaches

Diamond

Minimization of the model output quadratic error

Minimization of a distance between fuzzy numbers

Search for the model most appropriated with data

Tanaka

Minimization of the model output uncertainty

Optimization of a criterion under constraints

Search for the less uncertain model respecting the constraints

Which constraints?

Relationship between observed and predicted outputs:

- **necessity** model: predicted outputs included in observed ones
- **conjunction** model: no empty intersection between predicted and observed outputs
- **possibility** model: observed outputs included in predicted ones

⇒ data total uncertainty is taken in care
Fuzzy linear regression: possibilistic approach

Context of the study

- observed outputs: symmetrical triangular fuzzy intervals
  ⇒ identified model with triangular parameters
- study leaded for the identification of single input models:
  \[ \hat{Y}(x) = A_0 \oplus A_1 x \]
  with \( \oplus \) sum of fuzzy intervals

✓ concepts can be extended to the multi inputs case

- possibility model: observed inputs included in predicted ones
Fuzzy linear regression: possibilistic approach

Basic concepts

- use of $\alpha$-cuts
  - conventional intervals are handled
- minimization of a linear criterion under constraints:
  - criterion: a representation of the model uncertainty
  - constraints: inclusion of conventional intervals at the level $\alpha$

What are the limits of this approach?
The use of $\alpha$-cuts

- inclusion constraints defined for this level $\alpha$:
  
  \[
  [Y_j]_\alpha \subseteq [\hat{Y}_j]_\alpha
  \]

  \[\iff \begin{cases} 
  K\hat{Y}_j + (1 - \alpha)R\hat{Y}_j \geq KY_j + (1 - \alpha)RY_j \\
  K\hat{Y}_j - (1 - \alpha)R\hat{Y}_j \leq KY_j - (1 - \alpha)RY_j
  \end{cases} \]

- after the identification: parameters considered as valid $\forall \alpha \in [0, 1]$
- BUT: triangular identified model, *equality of the Kernels necessary* for the total inclusion
- Total inclusion $\forall \alpha \in [0, 1]$ not guaranteed!
Conventionnal criteria

- Minimization of the model uncertainty
  - How can it be quantified?
- Most used criterion (Tanaka): sum of the predicted intervals radius

\[
\text{Somme} = M \cdot R_{A0} + R_{A1} \cdot \sum_{j=1}^{M} | x_j |
\]

- BUT: minimization at the observed points
  - strongly dependant on the learning points: weak robustness
Fuzzy linear regression: possibilistic approach

Model representativity

Conventional linear model: study of the output variation on the domain

- kernel: \( M(\hat{Y}(x)) = K_{\hat{Y}(x)} = K_{A_0} + K_{A_1} \cdot x \)
- variation according to the sign of \( K_{A_1} \rightarrow \text{any variation} \)
- radius: \( R(\hat{Y}(x)) = R_{\hat{Y}(x)} = R_{A_0} + R_{A_1} \cdot |x| \)
- Radius always positive!
- Variation of the output radius limited by the input sign
Points to improve in the previous method: summary

- Is it possible to consider a model respecting the inclusion $\forall \alpha \in [0, 1]$?
- Is it possible to identify such a model with a more robust criterion?
- Is it possible to improve the representativity of a fuzzy linear model?
Propositions

Solution to the inclusion problem

- identification of a trapezoidal fuzzy model

→ inclusion constraints at two levels \( \alpha : \alpha = 0 \) and \( \alpha = 1 \)
  - \( \alpha = 1 : K_{Y_j} \in [K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+] \)
  - \( \alpha = 0 : [K_{Y_j} - R_{Y_j}, K_{Y_j} + R_{Y_j}] \subseteq [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+] \)

→ linear membership function

⇒ inclusion guaranteed \( \forall \alpha \in [0, 1] \)

Remark : model output for a data \( j \):

\[
\hat{Y}_j = ([K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+], [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+])
\]
Propositions

A more robust criterion

- **Objective**: make the criterion independent of data

→ definition of a total uncertainty on the domain $D$

→ vertical dimension (trapezes) taken into account

⇒ Minimization of $\text{Volume}$ delimited by the model on $D$

$$
\text{Volume} = R(K_{A_0}) + R(S_{A_0}) + (R(K_{A_1}) + R(S_{A_1})) \cdot |M(D)|
$$

→ independent of observed inputs $x_j$

→ still linear criterion
Propositions

A model more representative of the data

- To have any kind of the output radius variation:
  - the input variable sign must be modified
  - the linear behavior must be kept

⇒ The origin of the model is set on a bound of $D = [x_{\text{min}}, x_{\text{max}}]$

⇒ New model defined on the domain $D$:

$$\hat{Y}(x) = A_0 \oplus A_1(x - \text{shift})$$

Study of the variation of this new model output radius

$$R([S_{\hat{Y}}]) = R([S_{A_0}]) + R([S_{A_1}]) \cdot |x - \text{shift}|$$

→ $x - \text{shift} \geq 0 \ \forall x \in D$ ⇒ increasing radius

→ $x - \text{shift} \leq 0 \ \forall x \in D$ ⇒ decreasing radius

<table>
<thead>
<tr>
<th>output spread variation</th>
<th>↑</th>
<th>↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used model</td>
<td>$A_0 \oplus A_1(x - x_{\text{min}})$</td>
<td>$A_0 \oplus A_1(x - x_{\text{max}})$</td>
</tr>
</tbody>
</table>
Main steps
- What is the global tendency of the observed outputs radius on $D$? 
  ⇒ Choice of the appropriated value of shift
- Identification of trapezoidal coefficients:
  ⇒ Minimization of the criterion **Volume** under the inclusion constraints:
  - at $\alpha = 0$ and $\alpha = 1$ → total inclusion guaranteed

Optimization linear program

$$\min R(K_{A_0}) + R(S_{A_0}) + (R(K_{A_1}) + R(S_{A_1})) \cdot |M(D)|$$

s.t. \[ \begin{align*}
K_{Y_j} \in [K_{\hat{Y}_j}^-, K_{\hat{Y}_j}^+]
\end{align*} \]

\[ \begin{align*}
[K_{Y_j} - R_{Y_j}, K_{Y_j} + R_{Y_j}] \subseteq [S_{\hat{Y}_j}^-, S_{\hat{Y}_j}^+]
\end{align*} \]
Performance indicators

Comparison of two models: it is necessary to use performance indicators.

Two approaches:

- Indicators about the model uncertainty:
  - \textit{Sum} of the radius
  - \textit{Volume} delimited by the model on $D$

- Indicators about the model fitting with data:
  - \textit{Compatibility} between observed and predicted data: inclusion degree
  - \textit{Distance} between observed and predicted data: quadratic error
Extension to piecewise linear models

Global model:
- shifted submodels with trapezoidal coefficients
- each submodel defined on its own domain
⇒ independant submodels for identification and using

Identification method:
- data segmentation:
  - on observed kernels
  - on observed radius
⇒ More representative models!
- optimization of the parameters with the criterion Volume
⇒ identification of a global model with a minimal uncertainty and respecting the total inclusion
Generalization to multi-inputs models

Previous concepts applicable to multi-inputs \((x_i, i = 1, \ldots, N)\)

- Trapezoidal coefficients
- Appropriated shift for each \(x_i\)
- Criterion **Volume** can be extended, still linear

\[ \Rightarrow \text{identification of the optimal model for the total output uncertainty on the domain} \]

Case of a piecewise model

- Data segmentation : cartesian product
- Identification of submodels

\[ \Rightarrow \text{Independant submodels on each region} \]
Numerical examples: inclusion problem

8 observed data, \( D = [0.1, 0.8] \)
crisp inputs
observed outputs: symmetrical triangular fuzzy numbers
form of the identified model:
\[
\hat{Y}(x) = A_0 \oplus A_1 x
\]

<table>
<thead>
<tr>
<th>( j )</th>
<th>( x_j )</th>
<th>( Y_j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>(2.25, 0.75)</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>(2.875, 0.875)</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>(2.5, 1)</td>
</tr>
<tr>
<td>4</td>
<td>0.4</td>
<td>(4.25, 1.75)</td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>(4.0, 1.5)</td>
</tr>
<tr>
<td>6</td>
<td>0.6</td>
<td>(5.25, 1.25)</td>
</tr>
<tr>
<td>7</td>
<td>0.7</td>
<td>(7.5, 2)</td>
</tr>
<tr>
<td>8</td>
<td>0.8</td>
<td>(8.5, 1.5)</td>
</tr>
</tbody>
</table>
Numerical examples: inclusion problem

Triangular coefficients: identification at $\alpha = 0$

study of the inclusion: data $j = 1$

- Inclusion is respected at $\alpha = 0$
- Inclusion is not guaranteed $\forall \alpha$

<table>
<thead>
<tr>
<th>$A_0$</th>
<th>$A_1$</th>
<th>Compatabilité</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.96, 0.96)</td>
<td>(7.92, 2.92)</td>
<td>0.83</td>
</tr>
</tbody>
</table>
Numerical examples: inclusion problem

Trapezoidal coefficients

Study of the inclusion: data $j = 1$

Inclusion is respected $\forall \alpha$!
Numerical examples: criterion robustness

Presentation of the data set

- data \( j = 8 \) duplicated three times, \( D = [0.1, 0.8] \)
- study of the influence of data redundancy
- trapezoidal identified model
- 2 minimized criteria: \( \text{Sum} \) and \( \text{Volume} \)

<table>
<thead>
<tr>
<th></th>
<th>Somme</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_0 )</td>
<td>([-0.46, 1.36], [-1.96, 1.92])</td>
<td>([0.25, 1.36], [0, 1.92])</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>([8.93, 8.93], [8.93, 10.83])</td>
<td>([7.5, 8.93], [5, 10.83])</td>
</tr>
<tr>
<td>distance</td>
<td>71.05</td>
<td>91.29</td>
</tr>
<tr>
<td>somme</td>
<td>37.08</td>
<td>38.42</td>
</tr>
<tr>
<td>volume</td>
<td>3.28</td>
<td>3.15</td>
</tr>
</tbody>
</table>

- initial model identified again for \( \text{Volume} \)
- Minimal total uncertainty
- the indicators \( \text{distance} \) and \( \text{sum} \) are not robust!
Numerical examples: output representativity

Presentation of the data set

- Initial data shifted to have negative inputs
  ⇒ Increasing outputs radius
- Trapezoidal identified models, for the minimum Volume:
  - Conventional model: \( \hat{Y}(x) = A_0 \oplus A_1 x \)
  - Shifted model: \( \hat{Y}(x) = A_0 \oplus A_1 (x - \text{shift}) \)

<table>
<thead>
<tr>
<th>( j )</th>
<th>( x_j )</th>
<th>( Y_j )</th>
<th>( \hat{Y}(x) = A_0 \oplus A_1 \cdot x )</th>
<th>( \hat{Y}(x) = A_0 \oplus A_1 (x - \text{shift}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.8</td>
<td>(2.25, 0.75)</td>
<td>([7.57, 9.39], [0.07, 11.29])</td>
<td>([1.25, 0.53])</td>
</tr>
<tr>
<td>2</td>
<td>-0.7</td>
<td>(2.875, 0.875)</td>
<td>8.93</td>
<td>([7.5, 8.93], [5.10, 8.3])</td>
</tr>
<tr>
<td>3</td>
<td>-0.6</td>
<td>(2.5, 1)</td>
<td>58.83</td>
<td>48.08</td>
</tr>
<tr>
<td>4</td>
<td>-0.5</td>
<td>(4.25, 1.75)</td>
<td>28.14</td>
<td>25.17</td>
</tr>
<tr>
<td>5</td>
<td>-0.4</td>
<td>(4.0, 1.5)</td>
<td>3.52</td>
<td>3.15</td>
</tr>
<tr>
<td>6</td>
<td>-0.3</td>
<td>(5.25, 1.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.2</td>
<td>(7.5, 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.1</td>
<td>(8.5, 1.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All the indicators are better for the shifted model!
Numerical examples: output representativity

<table>
<thead>
<tr>
<th></th>
<th>$\hat{Y}(x) = A_0 \oplus A_1 \cdot x$</th>
<th>$\hat{Y}(x) = A_0 \oplus A_1 (x - shift)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0$</td>
<td>([7.57, 9.39], [0.07, 11.29])</td>
<td>([1.25, 0.5], [0.5, 3])</td>
</tr>
<tr>
<td>$A_1$</td>
<td>8.93</td>
<td>(7.5, 8.93), [5, 10.83])</td>
</tr>
<tr>
<td>distance</td>
<td>58.83</td>
<td>48.08</td>
</tr>
<tr>
<td>somme</td>
<td>28.14</td>
<td>25.17</td>
</tr>
<tr>
<td>volume</td>
<td>3.52</td>
<td>3.15</td>
</tr>
</tbody>
</table>

Conventional model

Model with shifted input
Numerical examples: piecewise model

- Great number of data
- Various tendencies of kernel and radius variations, noise: segmentation
  \[ \Rightarrow \text{Identification of shifted trapezoidal multi-inputs submodels} \]
  \[ \Rightarrow \text{Identification of the optimal global model for the minimal uncertainty (Volume)} \]
Numerical examples: higher order model

- Initial data set
- 2-order identified model:

\[ \hat{Y}(x) = A_0 \oplus A_1 x \oplus A_2 x^2 \]
Numerical examples: multi-inputs model

- Two inputs → two appropriated values of shifts:
  - Increasing radius on $x_1$ and decreasing one on $x_2$
  - Identification of the trapezoidal model for **Volume**
  ⇒ ”Planes” with total inclusion of data
Numerical examples: multi-inputs piecewise model

- Various tendencies of kernel and radius variations: segmentation
  ⇒ Identification of shifted trapezoidal multi-inputs submodels
  ⇒ Identification of the optimal global model for the minimum uncertainty (Volume)
Conclusion

Propositions

- Identification of trapezoidal models: total inclusion is guaranteed
- Improvement of the model representativity
- Identification for the minimal total uncertainty: increased robustness
- Application of these concepts on piecewise linear regression problems
Conclusion

Future work

- Imprecise observed inputs
  - Uncertainty comes from:
    - fuzzy parameters
    - fuzzy inputs
  - Lack of representativity of conventional space \((X, Y)\)

\[\Rightarrow\] Representation of the model in the space \((\text{Mid}(X), \text{Rad}(X), Y)\)

\[\Rightarrow\] Identification of the model in this space
Conclusion

Questions ?…

END