Hierarchical Structures of System Modeling with Information Granules

Witold Pedrycz

Department of Electrical & Computer Engineering
University of Alberta, Edmonton, AB, Canada
and
Systems Research Institute
Polish Academy of Sciences, Warsaw, Poland

wpedrycz@ualberta.ca

May 16, 2015
Agenda

Fuzzy sets and fuzzy models: a perspective

Fundamentals of Granular Computing

Fuzzy rule-based models and their granular generalizations

Hierarchy of granular models and granular outliers

Experiments

Conclusions
Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

Fuzzy sets and fuzzy modeling

numeric representation of membership grades (numeric membership functions)

fuzzy sets (1965)
- numeric representation of membership grades (numeric membership functions)
- type-2 fuzzy sets
- interval-valued fuzzy sets

granular fuzzy sets

Fuzzy models and their generalizations

fuzzy models
- type-2 fuzzy models
- granular fuzzy models
Are Fuzzy Models Fuzzy?

Existing plethora of fuzzy models

In their architectures and design schemes, there is a prevailing, if not common, trend of viewing fuzzy models as numeric constructs

- Architectures involving fuzzy sets

however

- Design guided by numeric performance index (RMSE)

- Use of fuzzy models as predominantly numeric constructs (decoding, defuzzification), their manifestation at the numeric level
Fuzzy models: design process

- Data
- Fuzzy model
- Performance index
  - optimization
  - decoding
- optimization of fuzzy sets

Data
Fuzzy models: design process (2)

- Data
- Type
- Fuzzy model
- Performance index
- Type reduction and decoding
- Optimization of fuzzy sets of type-2; computing overhead
- Optimization
From fuzzy models to granular fuzzy models

- Performance
- Evaluating
- Granularity allocation
- Modeling
- Data

Performance indexes
Design strategy
Granular Computing: An Introduction
Information granules: entities composed of elements being drawn together on a basis of similarity, functional closeness, temporal resemblance, spatial neighborhood, etc. and subsequently regarded as a single semantically meaningful unit used in processing.
Information granularity

Our ability to conceptualize the world at different granularities and to switch among these granularities is fundamental to our intelligence and flexibility.

It enables us to map the complexities of the world around us into simple theories that are computationally tractable to reason in.

Information granularity: recognition and classification
Information granularity
Information granules: key features

- Information granules as generic mechanisms of abstraction
- Customized, user-centric and business-centric approach to problem description and problem solving
- Processing at the level of information granules optimized with respect to the specificity of the problem
Information granules: from their conceptualization to realization

Humans

Information granules

Implicit information granules
Various points of view (models)

Computer realizations

Explicit (operational) information granules
Fuzzy sets
Rough sets
Intervals (sets)
Shadowed sets
Probability functions
Information granules of higher type

**Information granule of type 2** – granule whose characterization (description) is another information granule (not a single number)

Examples: type -2 fuzzy sets, interval-valued fuzzy sets, probabilistic sets, uncertain probabilities…

Temperature

- low temperature (-10°C) = 0.7
- high temperature (35°C) = 1.0

low temperature (-10°C) = [0.6, 0.8]
high temperature (35°C) = [0.95, 1.00]
Information granules of higher order

Information granule of order 2 – granule defined over a space composed of information granules

Examples: order -2 fuzzy sets

Temperature space of information granules
\[ X = \{ \text{low temperature, medium temperature, high temperature} \} \]

comfortable weather defined in \( X \)

- comfortable weather (low) = 0.4
- comfortable weather (medium) = 1.0
- comfortable weather (high) = 0.7
Fuzzy rule-based models
Fuzzy rule-based models

Modular models composed of conditional “if-then” statements describing behavior of system

- if condition then conclusion

with the condition (and conclusion) parts formalized in terms of Information granules
Rules in the form

If \( x \) is \( B_i(x) \) then \( \tilde{y}_i \) is \( f_i(x) \), \( i = 1, 2, \ldots, c \)

Local linear models

\[
f_i(x) = w_i + a_i^T(x - \nu_i)
\]

Aggregation of local models

\[
\hat{y} = \sum_{i=1}^{c} B_i(x) f_i(x) = \sum_{i=1}^{c} B_i(x)[w_i + a_i^T(x - \nu_i)]
\]
Takagi-Sugeno (TS) fuzzy models

Fuzzy set $B^\sim$ of order-2 defined over the space of information granules $\{B_1, B_2, \ldots, B_c\}$

If $x$ is $B_i(x)$ then $\tilde{y}_i$ is $f_i(x)$, $i = 1, 2, \ldots, c$
Takagi-Sugeno (TS) fuzzy models: detailed computing

\[
\hat{y} = \sum_{i=1}^{c} B_i(x)f_i(x) = \sum_{i=1}^{c} B_i(x)[w_i + a_i^T(x - \nu_i)]
\]

Introduce notation

\[
z_i = x - \nu_i \quad q = \sum_{i=1}^{c} B_i(x)w_i
\]

\[
\hat{y} = q + \sum_{i=1}^{c} a_i^T z_i
\]
Design process

Two key design phases:

Construction of condition parts of rules ($B_i$)

Determination of parameters of local linear functions ($a_i$)

Data $(x_1, y_1), (x_2, y_2)\ldots (x_N, y_N)$, $x_k$ in $\mathbb{R}^d$
Design process (1)

Construction of condition parts of rules ($B_i$)

Determination of structure in the $(d+1)$-dimensional input – output space $\mathbb{R}^{d+1}$

Fuzzy clustering (e.g., FCM) used in the development of $B_i$, $i=1, 2, \ldots, c$. The results are prototypes $[v_i, w_i]$ formed in $\mathbb{R}^{d+1}$ and clusters $B_i$ (condition parts) in the input space

Unsupervised learning – structure determination
Optimization of parameters of local linear models

Supervised learning with the performance index

\[
Q = \sum_{k=1}^{N} (y_k - \hat{y}_k)^2
\]

Minimize \( Q \) with respect to \( a_1, a_2, \ldots, a_c \)
\[ \hat{y}_k = q_k + \sum_{i=1}^{c} a_i^T z_{ki} \]

\[ z_{ki} = x_k - \nu_i \]

\[ p = [y_1 - q_1, y_2 - q_2, \ldots, y_N - q_N]^T \]

\[ a = [a_{11}, a_{12}, \ldots, a_{1d}, a_{21}, a_{22}, \ldots, a_{2d}, a_{31}, \ldots, a_{c1}, a_{c2}, \ldots, a_{cd}]^T \]
Detailed computing (2)

\[ \tilde{Z} = \begin{bmatrix}
    z_{11} & z_{12} & \cdots & z_{1c} \\
    z_{21} & z_{22} & \cdots & z_{2c} \\
    \vdots & \vdots & \ddots & \vdots \\
    z_{N1} & z_{N2} & \cdots & z_{Nc}
\end{bmatrix} \]

\[ Q = \sum_{k=1}^{N} (y_k - q_k - \sum_{i=1}^{c} a_i^T z_{ki})^2 = (p - \tilde{Z}a)^T (p - \tilde{Z}a) \]

\[ a_{opt} = (\tilde{Z}^T \tilde{Z})^{-1} \tilde{Z}^T p \]
Allocation of information granularity

- form granular parameters of the model on a basis of numeric models

- information granularity as a design asset
Allocation of information
granularity: protocols

\[ a_{ij} = \begin{cases} \min \left( a_{ij} (1 - \frac{e}{2}), a_{ij} (1 + \frac{e}{2}) \right) & \text{if } a_{ij} \neq 0 \\ \frac{e}{2} & \text{if } a_{ij} = 0 \end{cases} \]

\[ a_{ij}^+ = \begin{cases} \max \left( a_{ij} (1 - \frac{e}{2}), a_{ij} (1 + \frac{e}{2}) \right) & \text{if } a_{ij} \neq 0 \\ \frac{e}{2} & \text{if } a_{ij} = 0 \end{cases} \]
Allocation of information granularity: protocols

\[ a_{ij} = \begin{cases} \min(a_{ij}(1 - g_{ij}), a_{ij}(1 + (1 - g_{ij}) e)) & \text{if } a_{ij} > 0 \\ \max(a_{ij}(1 - g_{ij}), a_{ij}(1 + (1 - g_{ij}) e)) & \text{if } a_{ij} = 0 \end{cases} \]

\[ a_{ij}^+ = \begin{cases} \min(a_{ij}(1 - g_{ij}), a_{ij}(1 + (1 - g_{ij}) e)) & \text{if } a_{ij} > 0 \\ \max(a_{ij}(1 - g_{ij}), a_{ij}(1 + (1 - g_{ij}) e)) & \text{if } a_{ij} = 0 \end{cases} \]
Allocation of information granularity: protocols

\[ a_{ij} = \begin{cases} \min \left( a_{ij} (1 - g e), a_{ij} (1 + (1 - g) e) \right) & \text{if } a_{ij} \neq 0 \\ 0 & \text{if } a_{ij} = 0 \end{cases} \]

\[ a_{ij}^+ = \begin{cases} \max \left( a_{ij} (1 - g e), a_{ij} (1 + (1 - g) e) \right) & \text{if } a_{ij} \neq 0 \\ 0 & \text{if } a_{ij} = 0 \end{cases} \]
Allocation of information granularity: summary

(i) \( \frac{\varepsilon}{2} \quad \frac{\varepsilon}{2} \quad \frac{\varepsilon}{2} \quad \frac{\varepsilon}{2} \quad \ldots \quad \frac{\varepsilon}{2} \quad \frac{\varepsilon}{2} \)

(ii) \( \varepsilon \gamma \quad \varepsilon (1 - \gamma) \quad \varepsilon \gamma \quad \varepsilon (1 - \gamma) \quad \ldots \quad \varepsilon \gamma \quad \varepsilon (1 - \gamma) \)

(iii) \( \varepsilon \gamma_1 \quad \varepsilon (1 - \gamma_1) \quad \varepsilon \gamma_2 \quad \varepsilon (1 - \gamma_2) \quad \ldots \quad \varepsilon \gamma_n \quad \varepsilon (1 - \gamma_n) \)
Particle Swarm Optimization (PSO)

swarm of particles operating in a multidimensional search space

particle interacts with other particles and analyzes its own history:

update of velocity

\[ \mathbf{v}(\text{iter} + 1) = \mathbf{v}(\text{iter}) + c_1 r \cdot (\text{local best } \mathbf{x}(\text{iter})) + c_2 g \cdot (\text{global best } \mathbf{x}(\text{iter})) \]

- \( \xi \) - inertial weight, \( c_1 \) - cognitive factor \( c_2 \) – social factor
- **local-best** -- the best position of the particle so far
- **global-best** - the best position in the swarm so far
- \( r, g \) - random vectors coming from \( \text{U}[0,2] \)

New position

\[ \mathbf{x}(\text{iter} + 1) = \mathbf{x}(\text{iter}) + \mathbf{v}(\text{iter} + 1) \]
Granular fuzzy model

granular (interval-valued) parameters \([ a_i^-, a_i^+ ]\)

\[
\hat{y} = \sum_{i=1}^{c} B_i(x) f_i(x) = \sum_{i=1}^{c} B_i(x)[w_i + a_i^T (x - \nu_i)]
\]

\[
z_i = x - \nu_i
\]

\[
q = \sum_{i=1}^{c} B_i(x) w_i
\]

\[
[a, b] \square [c, d] = [a + c, b + d]
\]

\[
[a, b] \quad [c, d] = [\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)]
\]
Evaluation criteria

Coverage criterion

\[
\text{cov} = \frac{1}{N} \sum_{k=1}^{N} \text{incl}(y_k, Y_k)
\]

Specificity criterion

\[
\text{spec} = \frac{1}{N} \sum_{k=1}^{N} \exp\left( - \left| y_k^+ \ y_k \right| \right)
\]

Both criteria depends on the assumed value of \( \varepsilon \)
For given $\varepsilon$ maximize coverage and report specificity values

\[
\text{cov} = \frac{1}{N} \sum_{k=1}^{N} \text{incl} \left( y_k, Y_k \right)
\]

\[
\text{spec} = \frac{1}{N} \sum_{k=1}^{N} \exp \left( \left| y_k^+ - y_k \right| \right)
\]
Optimization: a general evaluation

Coverage and specificity depend on values of $\varepsilon$

Coordinates of coverage and specificity

Area under curve (AUC)

Performance evaluated in terms of AUC
Experiments-synthetic data

\[ y = (1 + x_1^2 + x_2^{1.5})^2 \]
Experiments - synthetic data

Synthetic data

\[ y = (1 + x_1^2 + x_2^{1.5})^2 \]
Experiments - synthetic data

Local linear models (c = 5, m = 2.0)
Experiments- synthetic data

Local linear models (c =9, m=2.0)
Experiments - synthetic data

Training data

Testing data
Experiments- synthetic data

Training data

Testing data
Experiments - synthetic data

- $c = 2$
- $c = 5$
- $c = 9$

- $m = 1.1$

- $m = 2.0$

- $m = 2.9$
Experiments: Machine Learning data sets

<table>
<thead>
<tr>
<th>Name of the data</th>
<th>Number of instances</th>
<th>Number of inputs</th>
<th>Origin of the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston housing</td>
<td>506</td>
<td>13</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td>Auto–MPG</td>
<td>392</td>
<td>7</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td>Stock</td>
<td>950</td>
<td>9</td>
<td>StatLib repository</td>
</tr>
<tr>
<td>Concrete Slump</td>
<td>103</td>
<td>9</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td>Yacht</td>
<td>308</td>
<td>6</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td>Hydrodynamics</td>
<td></td>
<td></td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td>Forest fires</td>
<td>517</td>
<td>12</td>
<td>UCI machine learning repository</td>
</tr>
</tbody>
</table>
Publicly available data sets used in experiments - a brief description

<table>
<thead>
<tr>
<th>Name of the data</th>
<th>Number of instances</th>
<th>Number of inputs</th>
<th>Origin of the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston housing</td>
<td>506</td>
<td>13</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><a href="https://archive.ics.uci.edu/ml/">https://archive.ics.uci.edu/ml/</a></td>
</tr>
<tr>
<td>Auto-MPG</td>
<td>392</td>
<td>7</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><a href="http://archive.ics.uci.edu/ml/">http://archive.ics.uci.edu/ml/</a></td>
</tr>
<tr>
<td>Stock</td>
<td>950</td>
<td>9</td>
<td>StatLib repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><a href="http://www.dcc.fc.up.pt/~ltorgo/Regression/">http://www.dcc.fc.up.pt/~ltorgo/Regression/</a></td>
</tr>
<tr>
<td>Concrete Slump</td>
<td>103</td>
<td>9</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><a href="http://archive.ics.uci.edu/ml/">http://archive.ics.uci.edu/ml/</a></td>
</tr>
<tr>
<td>Yacht</td>
<td>308</td>
<td>6</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td>Hydrodynamics</td>
<td></td>
<td></td>
<td><a href="http://archive.ics.uci.edu/ml/">http://archive.ics.uci.edu/ml/</a></td>
</tr>
<tr>
<td>Forest fires</td>
<td>517</td>
<td>12</td>
<td>UCI machine learning repository</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><a href="https://archive.ics.uci.edu/ml/">https://archive.ics.uci.edu/ml/</a></td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td>Testing</td>
<td>Training</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>Boston housing</strong></td>
<td>3.4892±0.0676</td>
<td>3.7605±0.5430</td>
<td>3.0078±0.0703</td>
</tr>
<tr>
<td><strong>Auto MPG</strong></td>
<td>2.7918±0.0540</td>
<td>2.8698±0.4809</td>
<td>2.6695±0.0496</td>
</tr>
<tr>
<td><strong>Stock</strong></td>
<td>1.4843±0.0134</td>
<td>1.4905±0.1256</td>
<td>1.0080±0.0234</td>
</tr>
<tr>
<td><strong>Concrete Slump</strong></td>
<td>1.9926±0.0652</td>
<td>2.4860±0.6233</td>
<td>1.2082±0.0857</td>
</tr>
<tr>
<td><strong>Yacht Hydrodynamics</strong></td>
<td>2.6702±0.0636</td>
<td>24.2685±2.2786</td>
<td>0.9445±0.0513</td>
</tr>
<tr>
<td><strong>Forest fires</strong></td>
<td>1.3312±0.0149</td>
<td>1.4235±0.1696</td>
<td>1.2883±0.0151</td>
</tr>
</tbody>
</table>
Experiments- Machine Learning data sets
Experiments- Machine Learning data sets

Performance of PSO for the third scenario of allocation of information granularity: (a) Boston housing \(c = 2\), (b) Auto MPG \(c = 9\), (c) Stock \(c = 2\), (d) Concrete Slump \(c = 5\), (e) Yacht Hydrodynamics \(c = 9\), (f) Forest fires \(c = 5\), (g) Boston housing \(c = 2\), (h) Stock \(c = 2\), (i) Concrete Slump \(c = 5\). Dashed lines: population - 10, solid lines: population - 20, Dash-dot lines: population - 40.
Experiments- Machine Learning data sets
Experiments- Machine Learning data sets

translucent bars - $c = 2$, gray bars - $c = 5$, and black bars - $c = 9$. 
Hierarchy of granular models
Granular fuzzy models of higher type

- **G^2-FM**
  - **Type 2**
  - Coverage/specificity criteria

- **G-FM**
  - **Type 1**
  - Coverage/specificity criteria

- **FM**
  - **Type 0**
Granular fuzzy models
granular intervals

Type 2

Type 1

Type 0
Granular fuzzy models
Interval-valued fuzzy sets
Granular fuzzy models
Granular outliers

outlier of type-2

outlier of type-1

outlier of type-0

G²-FM coverage/specificity criteria
G-FM coverage/specificity criteria
FM
Conclusions

Granular Computing as a general conceptual and algorithmic framework supporting design and analysis pursuits for system modeling and augmenting the methodology of fuzzy modeling

Fundamentals of Granular Computing

Plethora of further detailed algorithmic developments (with specific realizations of information granules – intervals/sets, rough sets, fuzzy sets…)