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

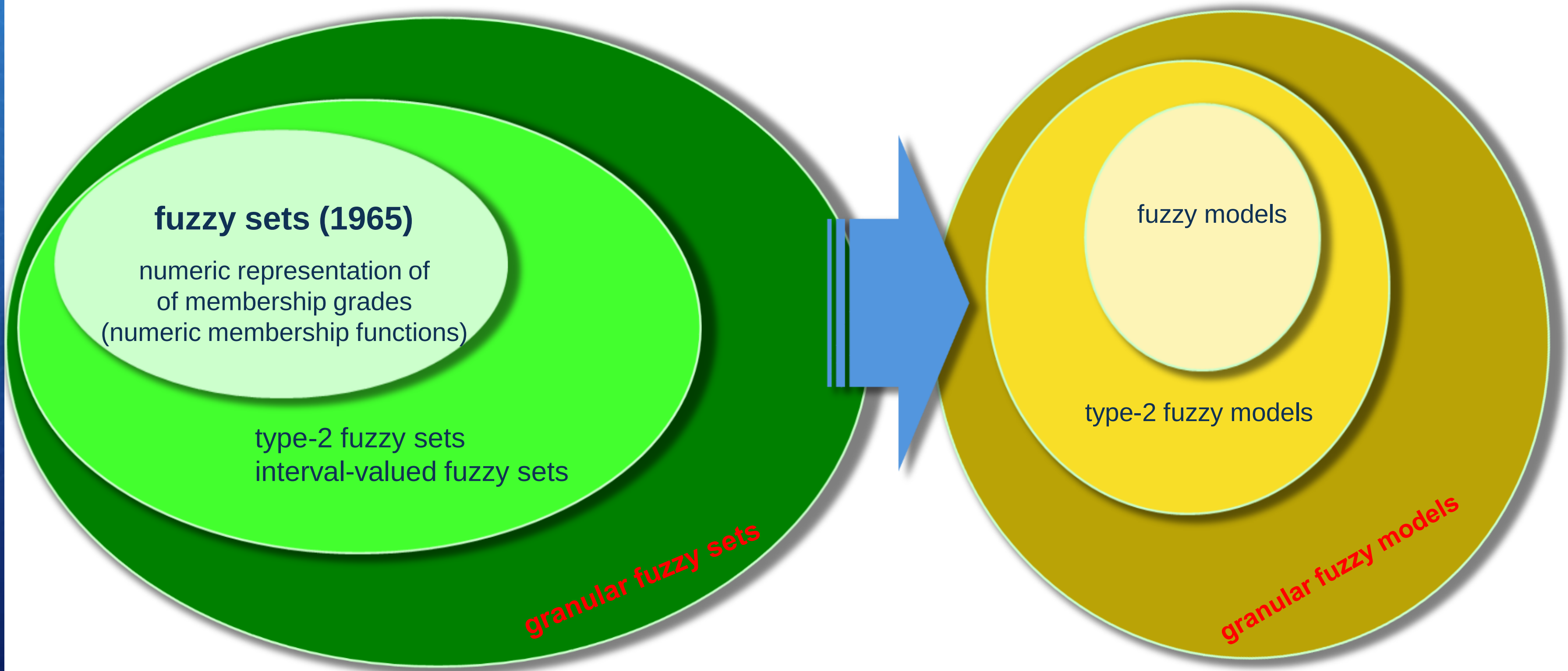
Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist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

J.E.P. Box Box, G. E. P. (1976), "Science and Statistics", Journal of the American Statistical Association ***71****: 791–799.1976*

Fuzzy sets and fuzzy modeling

numeric representation of
of membership grades
(numeric membership functions)

Fuzzy models and their
generalizations



Are Fuzzy Models Fuzzy ?

Existing plethora of fuzzy models

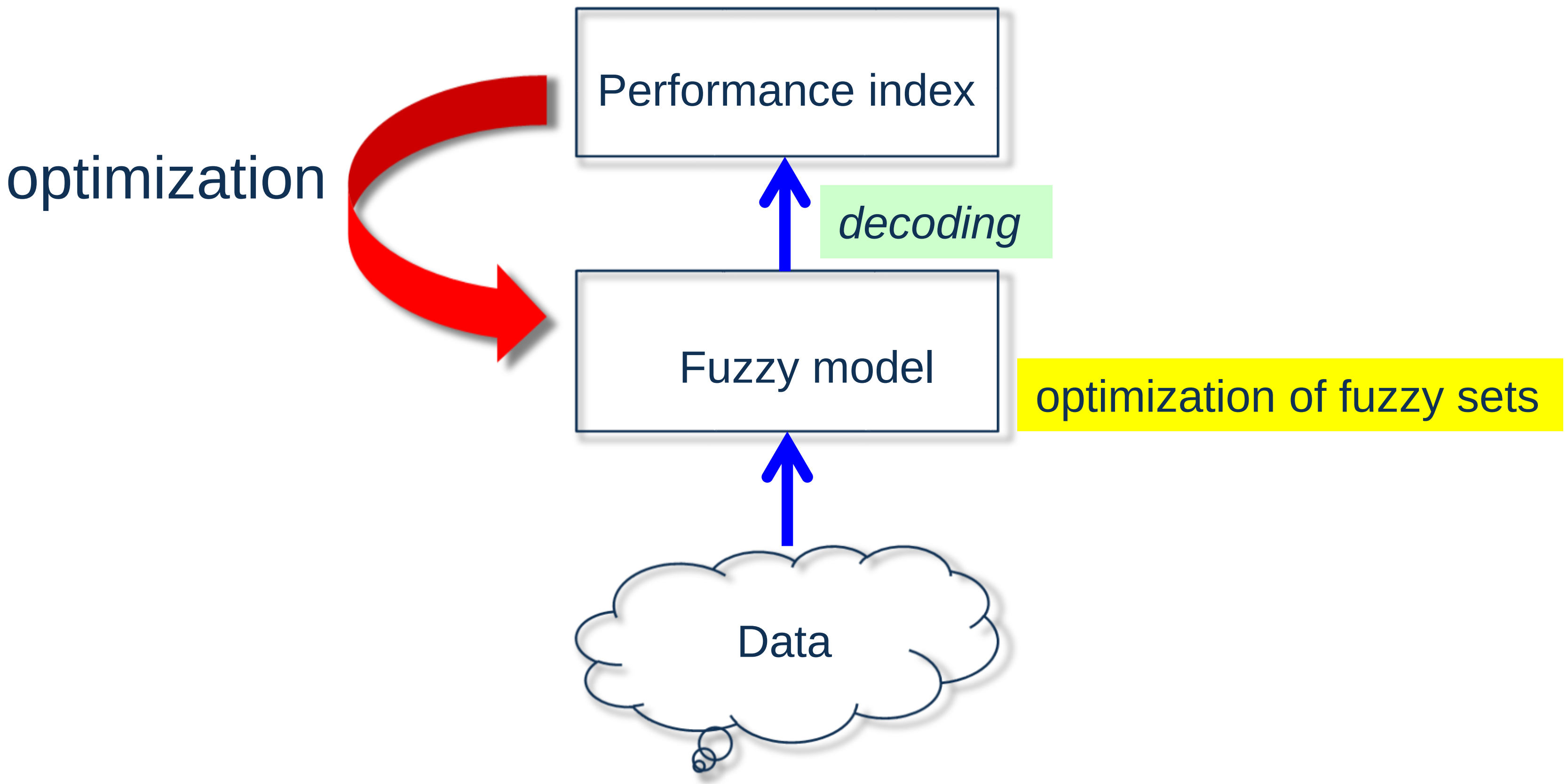
In their architectures and and design schemes, there is a prevailing, if not common, trend of viewing fuzzy models as numeric constructs

- Architectures involving fuzzy sets

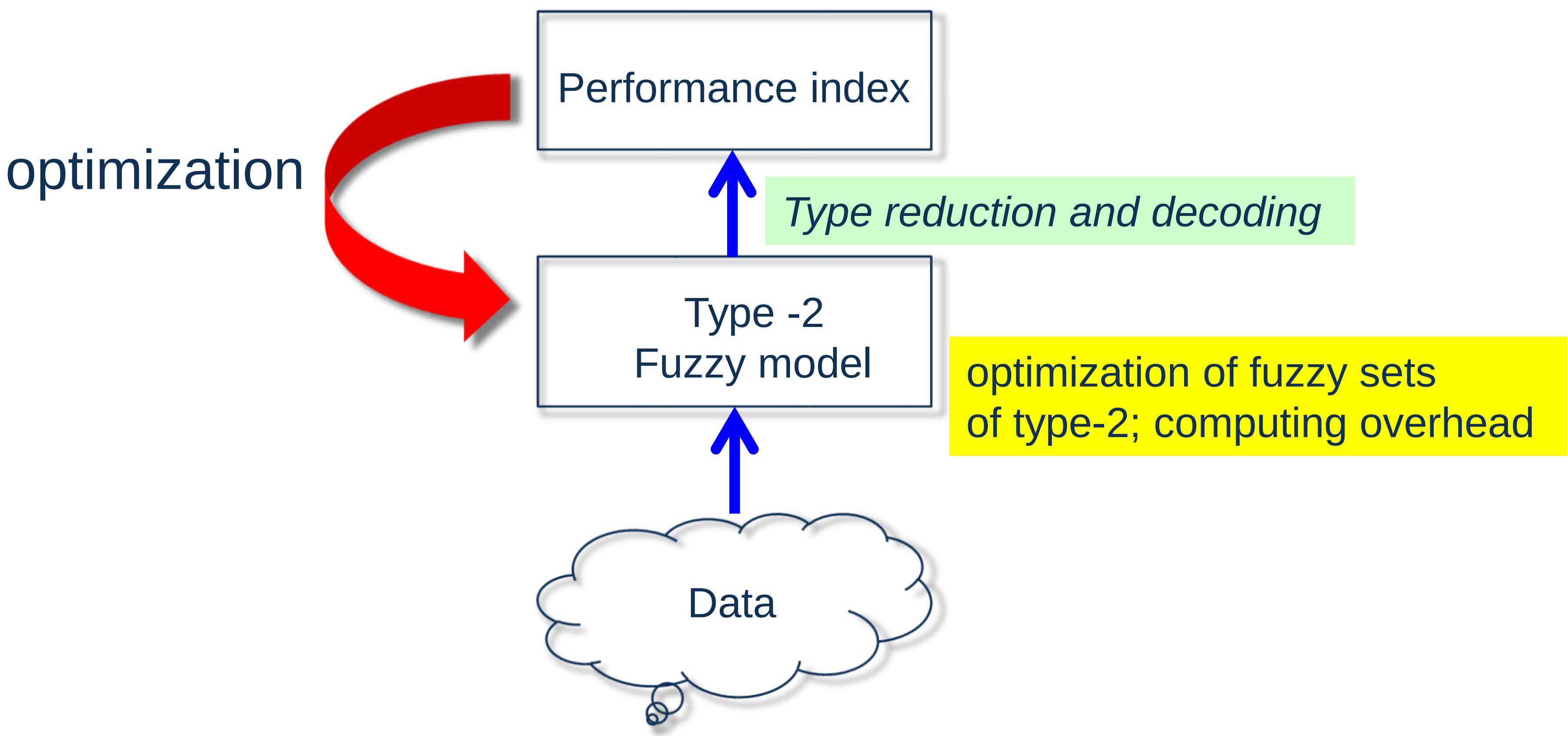
however

- Design guided by numeric performance index (RMSE)
- Use of fuzzy models as predominantly numeric constructs (decoding, defuzzification), their manifestation at the numeric level

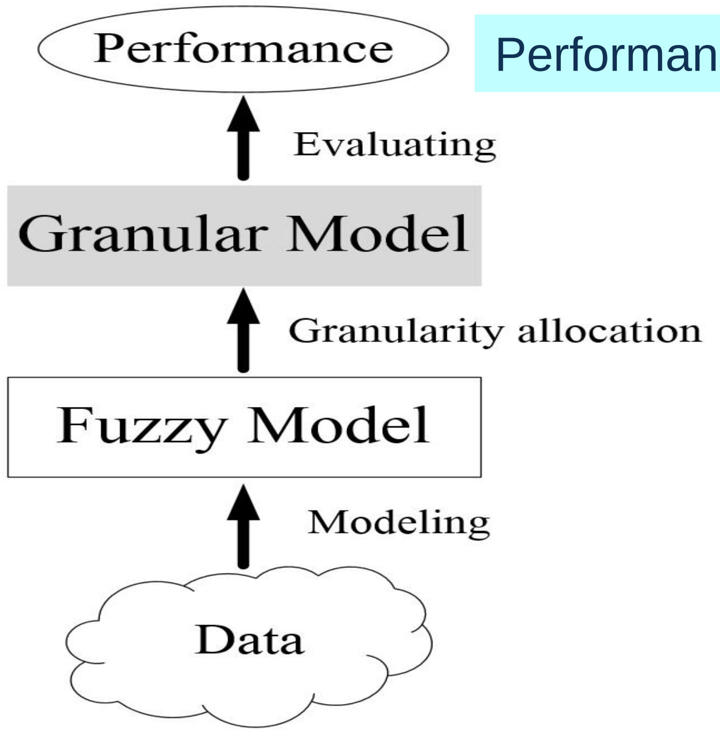
Fuzzy models: design process



Fuzzy models: design process(2)



From fuzzy models to granular fuzzy models



Performance indexes

Design strategy

Granular Computing: An Introduction

Information granules

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.

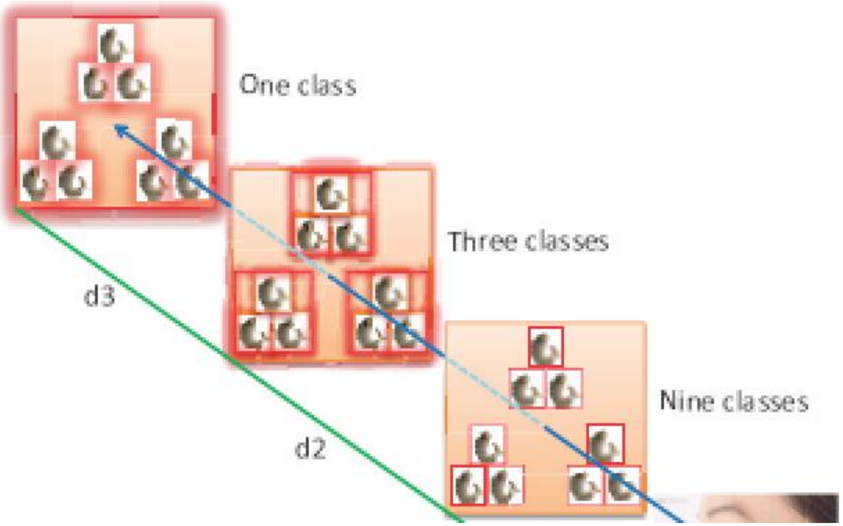
Information granularity

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

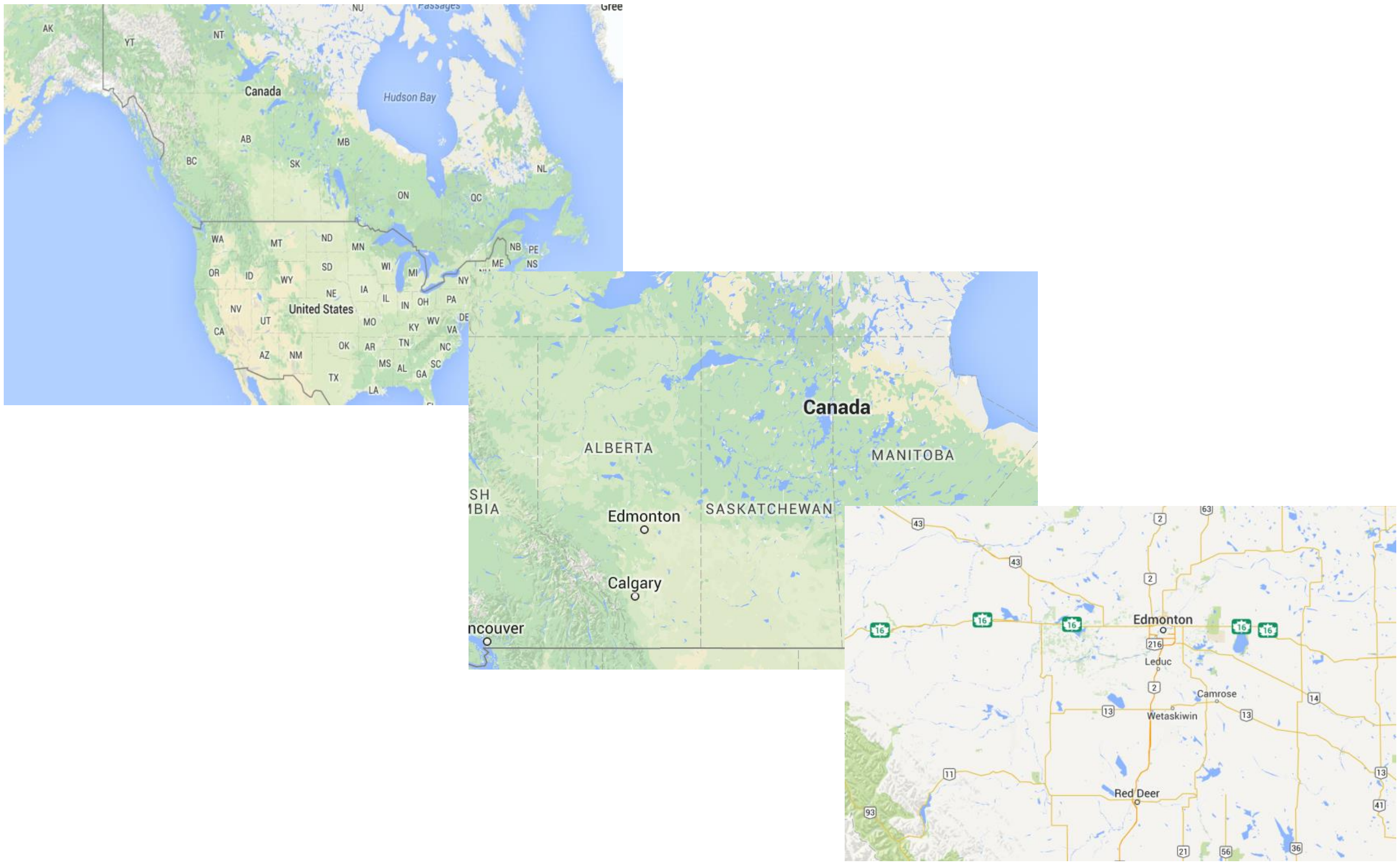
It enables us to map the complexities of the world around us into simple theories that are computationally tractable to reason in.

J. R. Hobbs, Proc. IJCAI, 1985

Information granularity: recognition and classification



Information granularity



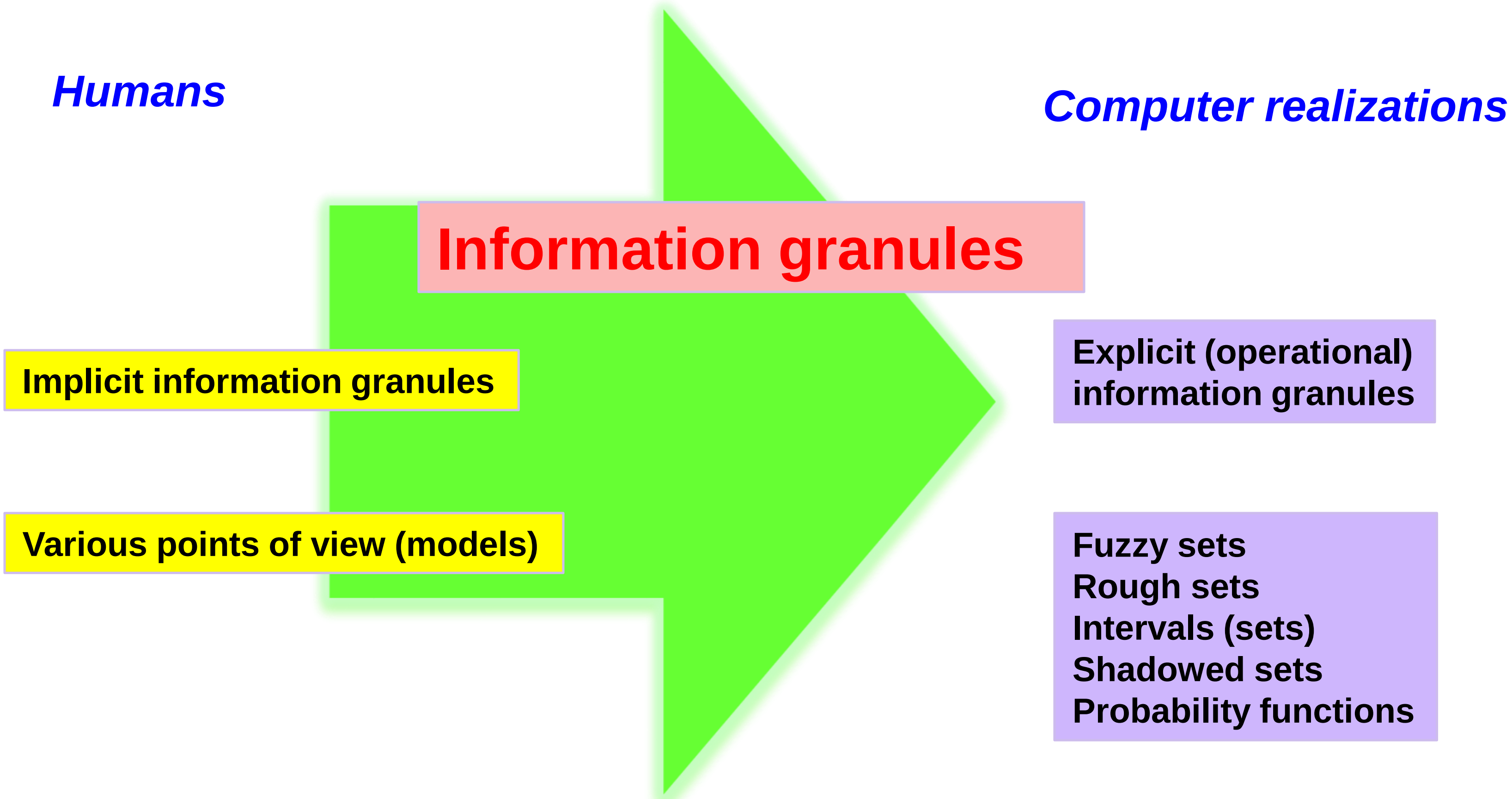
Information granules: key features

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



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

low temperature (-10C) =[0.6, 0.8]
high temperature (35C) =[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= {*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

Takagi-Sugeno (TS) fuzzy models

Rules in the form

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

Local linear models

$$f_i(\boldsymbol{x}) = w_i + \boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{v}_i)$$

Aggregation of local models

$$\hat{y} = \sum_{i=1}^{c} B_i(\boldsymbol{x})\mathrm{f}_i(\boldsymbol{x}) = \sum_{i=1}^{c} B_i(\boldsymbol{x})[w_i + \boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{v}_i)]$$

Takagi-Sugeno (TS) fuzzy models

B_1

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

B_c

$B^{\sim}$

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

Takagi-Sugeno (TS) fuzzy models: detailed computing

$$\hat{y} = \sum_{i=1}^{c} B_i(\boldsymbol{x})\,\mathrm{f}_i(\boldsymbol{x}) = \sum_{i=1}^{c} B_i(\boldsymbol{x})[w_i + \boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{v}_i)]$$

Introduce notation

$$\boldsymbol{z}_i = \boldsymbol{x} - \boldsymbol{v}_i \qquad q = \sum_{i=1}^{c} B_i(\boldsymbol{x}) w_i$$

$$\boxed{\hat{y} = q + \sum_{i=1}^{c} \boldsymbol{a}_i^T \boldsymbol{z}_i}$$

Design process

Two key design phases:

Construction of condition parts of rules (B_i)

Determination of parameters of local linear functions ($\mathbf{a}_i$)

Data $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \ldots (\mathbf{x}_N, y_N)$, x_k in $\mathbf{R}^d$

Design process (1)

Construction of condition parts of rules (B_i)

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

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

Unsupervised learning – structure determination

Design process (2)

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 $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_c$

Detailed computing (1)

$$\hat{y}_k = q_k + \sum_{i=1}^{c} \boldsymbol{a}_i^T \boldsymbol{z}_{ki}$$

$$\boldsymbol{z}_{ki} = \boldsymbol{x}_k - \boldsymbol{v}_i$$

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

$$\boldsymbol{a} = [a_{11}, a_{12}, \ldots, a_{1d}, a_{21}, a_{22}, \ldots, a_{2d}, a_{31} \ldots, a_{c1}, a_{c2}, \ldots, a_{cd}]^T$$

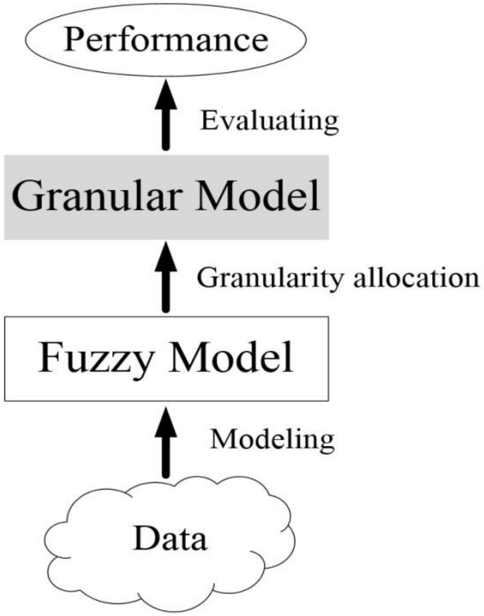
Detailed computing (2)

$$\tilde{Z} = \begin{bmatrix} \boldsymbol{z}_{11} & \boldsymbol{z}_{12} & \cdots & \boldsymbol{z}_{1c} \\ \boldsymbol{z}_{21} & \boldsymbol{z}_{22} & \cdots & \boldsymbol{z}_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{z}_{N1} & \boldsymbol{z}_{N2} & \cdots & \boldsymbol{z}_{Nc} \end{bmatrix}$$

$$Q = \sum_{k=1}^{N} (y_k - q_k - \sum_{i=1}^{c} \boldsymbol{a}_i^T \boldsymbol{z}_{ki})^2 = (\boldsymbol{p} - \tilde{Z}\boldsymbol{a})^T (\boldsymbol{p} - \tilde{Z}\boldsymbol{a})$$

$$\boldsymbol{a}_{\mathrm{opt}} = (\tilde{Z}^T \tilde{Z})^{-1} \tilde{Z}^T \boldsymbol{p}$$

Allocation of information granularity



Information granularity allocation

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

-information granularity as a design asset

Allocation of information granularity: protocols

$$a_{ij}^{-} = \begin{cases} \min\left(a_{ij}(1-\frac{\varepsilon}{2}), a_{ij}(1+\frac{\varepsilon}{2})\right) & \text{if } a_{ij} \neq 0 \\ -\varepsilon/2 & \text{if } a_{ij} = 0 \end{cases}$$

$$a_{ij}^{+} = \begin{cases} \max\left(a_{ij}(1-\frac{\varepsilon}{2}), a_{ij}(1+\frac{\varepsilon}{2})\right) & \text{if } a_{ij} \neq 0 \\ \varepsilon/2 & \text{if } a_{ij} = 0 \end{cases}$$

Allocation of information granularity: protocols

$$\alpha_{ij}^{-} = \begin{cases} \min\left(\alpha_{ij}(1-\gamma_{ij}\varepsilon), \alpha_{ij}(1+(1-\gamma_{ij})\varepsilon)\right) & \text{if } \alpha_{ij} \neq 0 \\ -\varepsilon\gamma_{ij} & \text{if } \alpha_{ij} = 0 \end{cases}$$

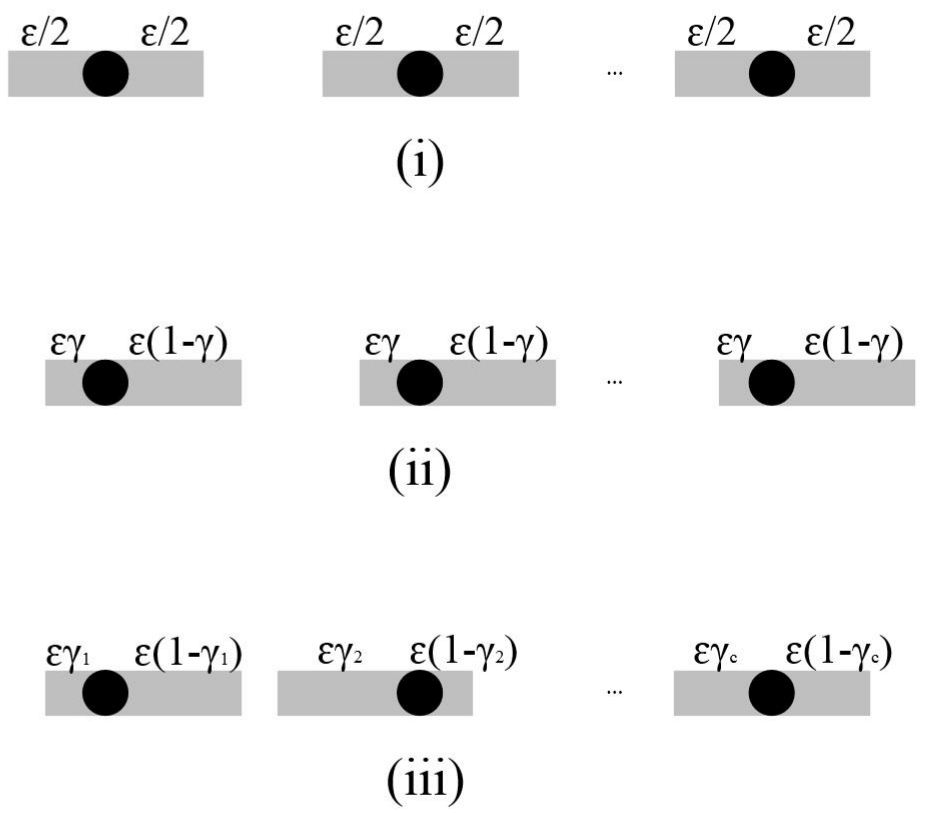
$$\alpha_{ij}^{+} = \begin{cases} \max\left(\alpha_{ij}(1-\gamma_{ij}\varepsilon), \alpha_{ij}(1+(1-\gamma_{ij})\varepsilon)\right) & \text{if } \alpha_{ij} \neq 0 \\ \varepsilon\gamma_{ij} & \text{if } \alpha_{ij} = 0 \end{cases}$$

Allocation of information granularity: protocols

$$\alpha_{ij}^{-} = \begin{cases} \min\left(\alpha_{ij}(1-\gamma\varepsilon), \alpha_{ij}(1+(1-\gamma)\varepsilon)\right) & \text{if } \alpha_{ij} \neq 0 \\ -\varepsilon\gamma & \text{if } \alpha_{ij} = 0 \end{cases}$$

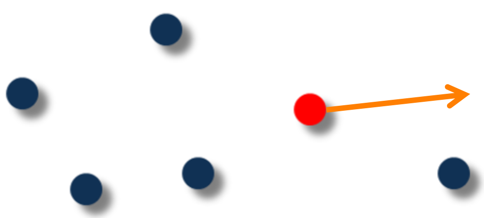
$$\alpha_{ij}^{+} = \begin{cases} \max\left(\alpha_{ij}(1-\gamma\varepsilon), \alpha_{ij}(1+(1-\gamma)\varepsilon)\right) & \text{if } \alpha_{ij} \neq 0 \\ \varepsilon\gamma & \text{if } \alpha_{ij} = 0 \end{cases}$$

Allocation of information granularity: summary



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

$$\mathbf{v}(\mathrm{iter}+1) = \mathbf{v}(\mathrm{iter}) + c_1 \mathbf{r} \cdot (\mathbf{local-best} - \mathbf{x}(\mathrm{iter})) + c_2 \mathbf{g} \cdot (\mathbf{global-best} - \mathbf{x}(\mathrm{iter}))$$

ξ- 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 U[0,2]

New position $$\mathbf{x}(\mathrm{iter}+1) = \mathbf{x}(\mathrm{iter}) + \mathbf{v}(\mathrm{iter}+1)$$

Granular fuzzy model

granular (interval-valued) parameters [$\boldsymbol{a}_i^-$, $\boldsymbol{a}_i^+$]

$$\hat{y} = \sum_{i=1}^{c} B_i(\boldsymbol{x})\,\mathrm{f}_i(\boldsymbol{x}) = \sum_{i=1}^{c} B_i(\boldsymbol{x})[w_i + \boldsymbol{a}_i^T(\boldsymbol{x} - \boldsymbol{v}_i)]$$

$$\boldsymbol{z}_i = \boldsymbol{x} - \boldsymbol{v}_i$$

$$q = \sum_{i=1}^{c} B_i(\boldsymbol{x}) w_i$$

$$\hat{y} = q \oplus \sum_{\oplus i=1}^{c} ([\boldsymbol{a}_i^-, \boldsymbol{a}_i^+]^T \otimes \boldsymbol{z}_i)$$

[a,b] □ [c,d] = [a+c, b+d]

[a,b] Ä [c,d] = [min(ac, ad, bc, bd), max(ac, ad, bc, bd)]

Evaluation criteria

Coverage criterion

$$cov = \frac{1}{N}\sum_{k=1}^{N} \mathrm{incl}\left(y_k, Y_k\right)$$

Specificity criterion

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

Both criteria depends on the assumed value of ε

Optimization

For given ε maximize coverage and report specificity values

$$cov = \frac{1}{N}\sum_{k=1}^{N} \mathrm{incl}\left(y_k, Y_k\right)$$

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

Optimization: a general evaluation

Coverage and specificity depend on values of ε

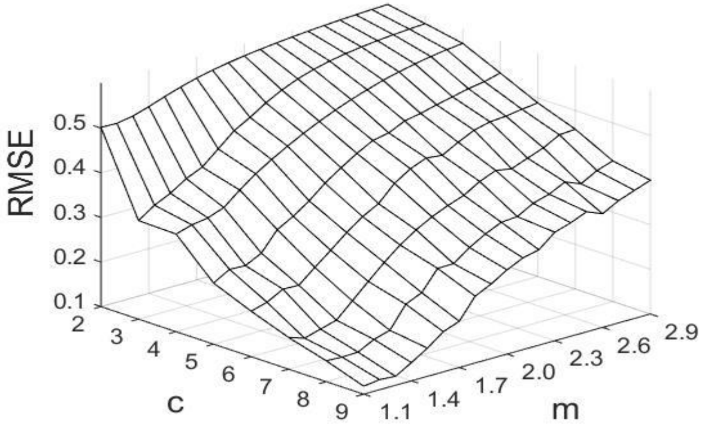
Coordinates of coverage and specificity

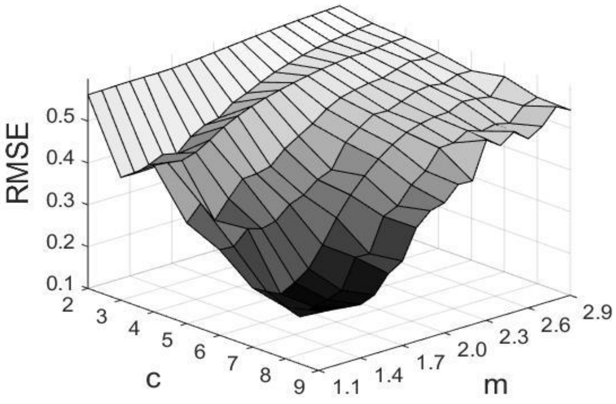
Area under curve (AUC)

Performance evaluated in terms of AUC

Experiments-synthetic data

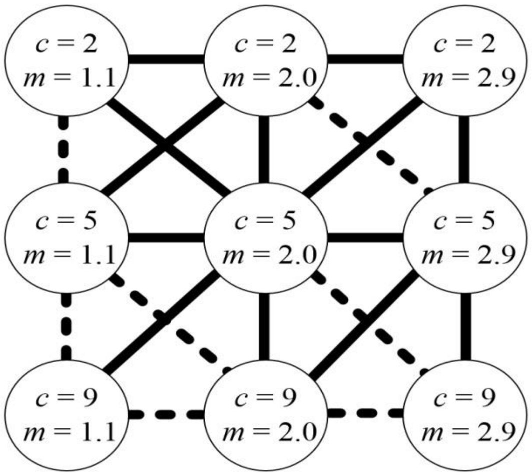
$$y = (1 + x_1^{-2} + x_2^{-1.5})^2$$





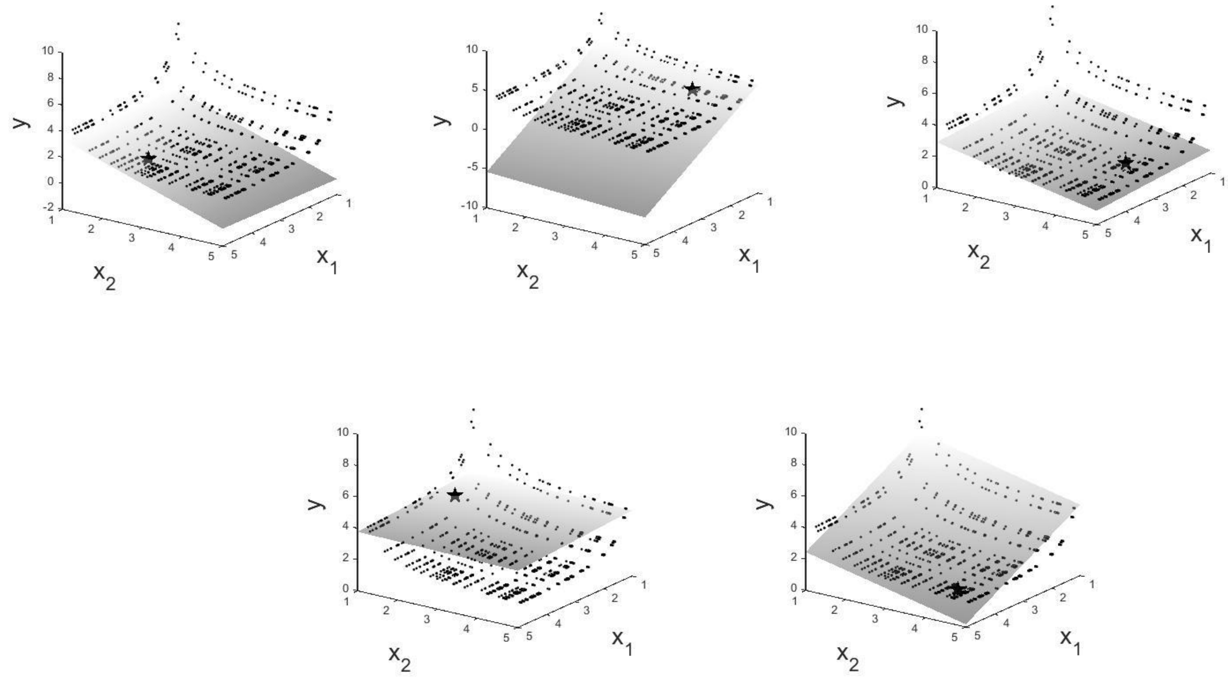
Experiments- synthetic data

Synthetic data $y = (1 + x_1^{-2} + x_2^{-1.5})^2$



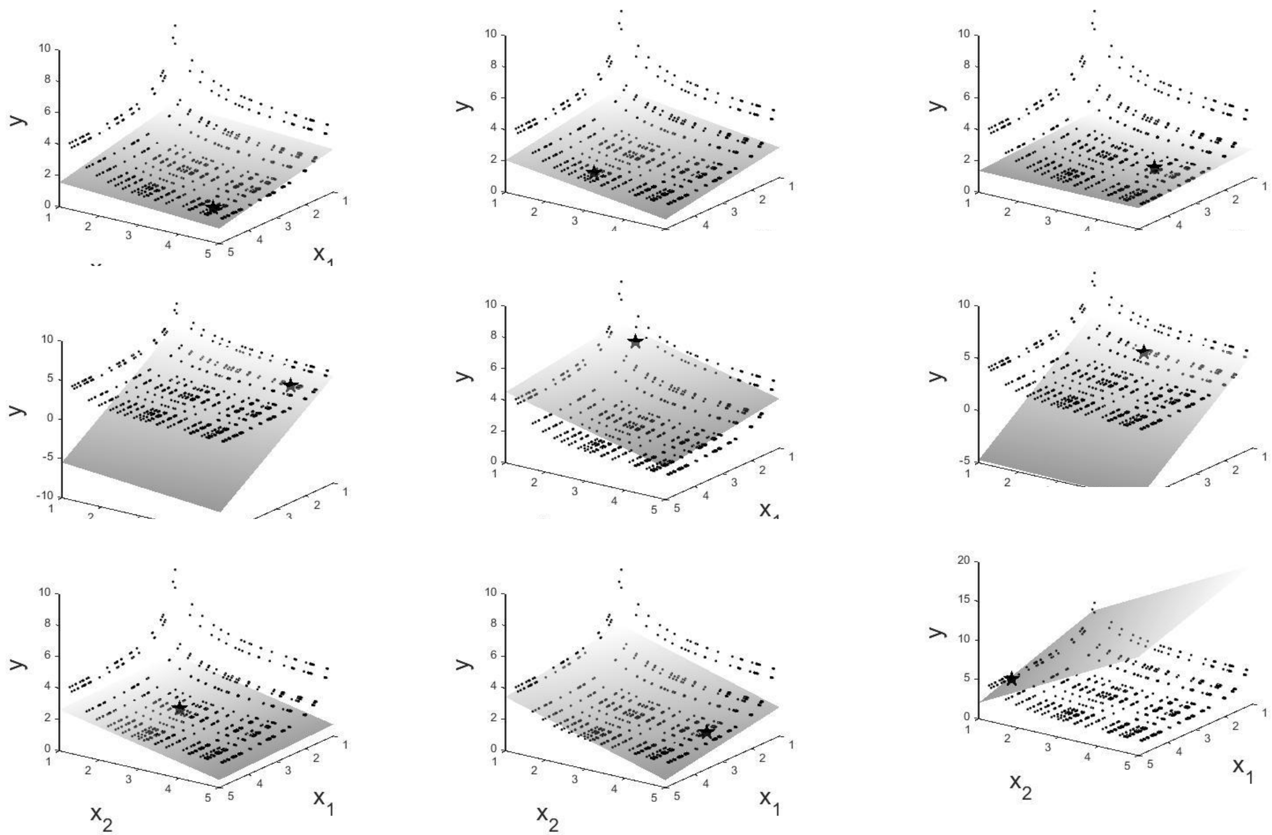
Experiments- synthetic data

Local linear models (c =5, m=2.0)

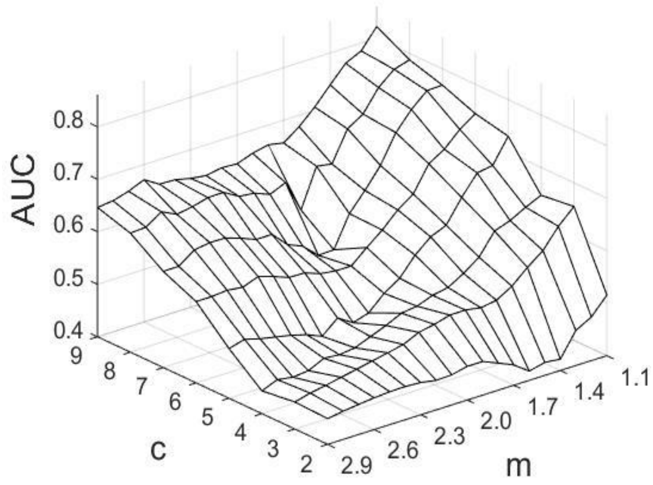


Experiments- synthetic data

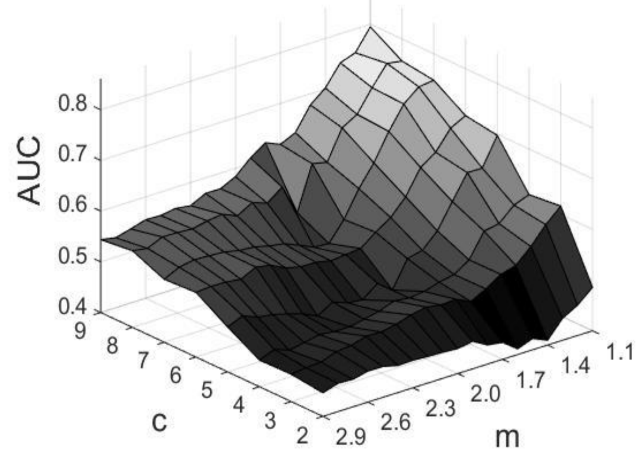
Local linear models (c =9, m=2.0)



Experiments- synthetic data

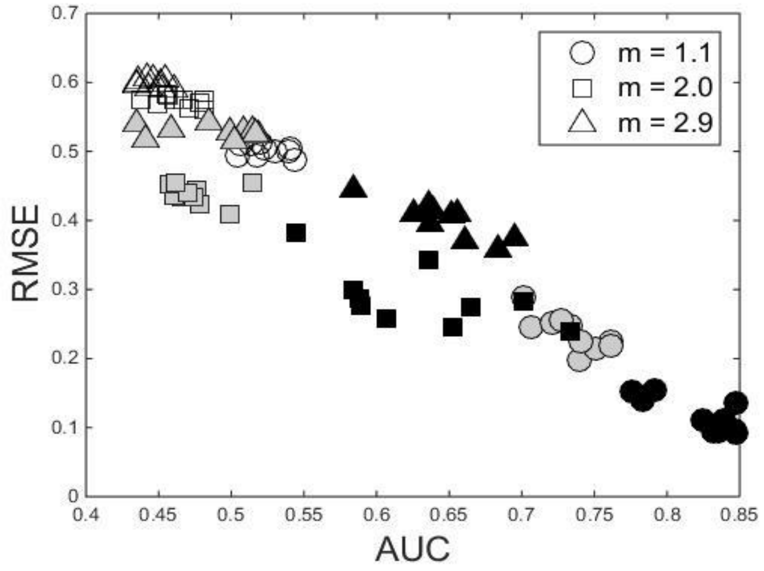


Training data

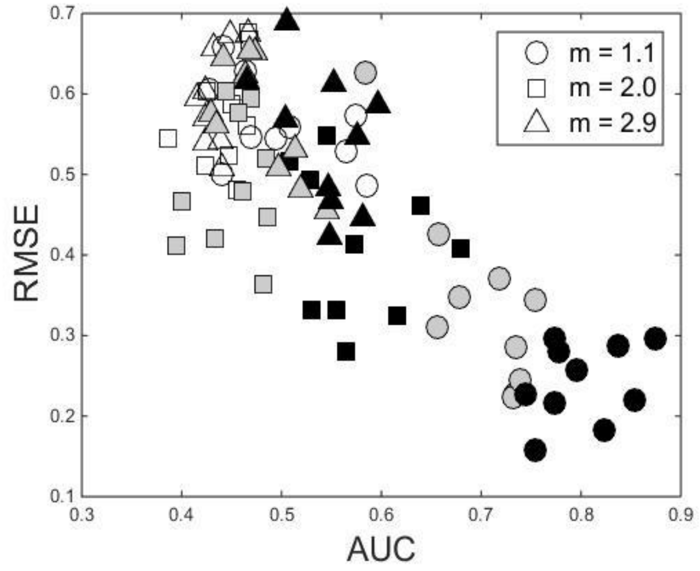


Testing data

Experiments- synthetic data



Training data



Testing data

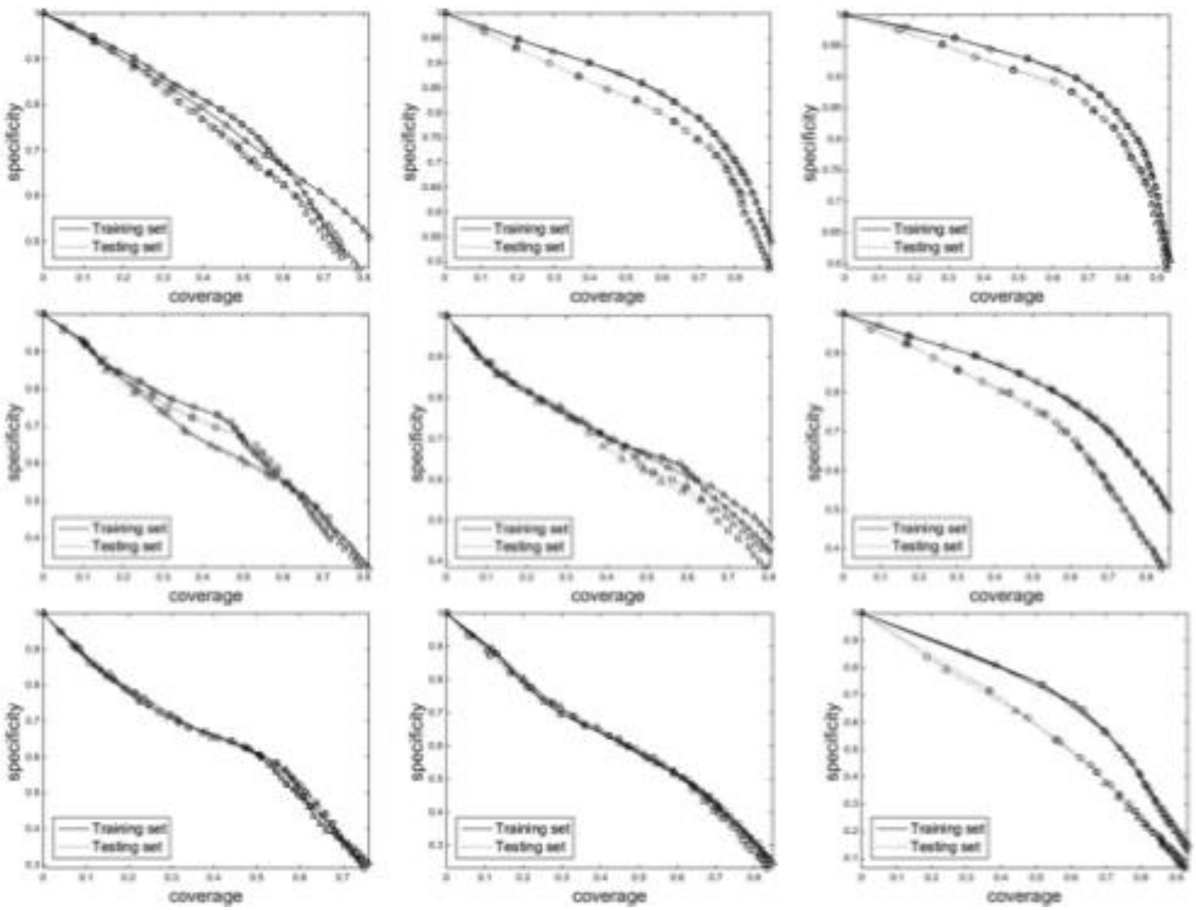
Experiments- synthetic data

c =2 c =5 c =9

m =1.1

m =2.0

m =2.9



Experiments- Machine Learning data sets

Publicly available data sets used in experiments- a brief description

Name of the data	Number of instances	Number of inputs	Origin of the data
Boston housing	506	13	UCI machine learning repository https://archive.ics.uci.edu/ml/
Auto–MPG	392	7	UCI machine learning repository http://archive.ics.uci.edu/ml/
Stock	950	9	StatLib repository http://www.dcc.fc.up.pt/~ltorgo/Regression/
Concrete Slump	103	9	UCI machine learning repository http://archive.ics.uci.edu/ml/
Yacht Hydrodynamics	308	6	UCI machine learning repository http://archive.ics.uci.edu/ml/
Forest fires	517	12	UCI machine learning repository https://archive.ics.uci.edu/ml/

$m=2$

Experiments- Machine Learning data sets

Publicly available data sets used in experiments- a brief description

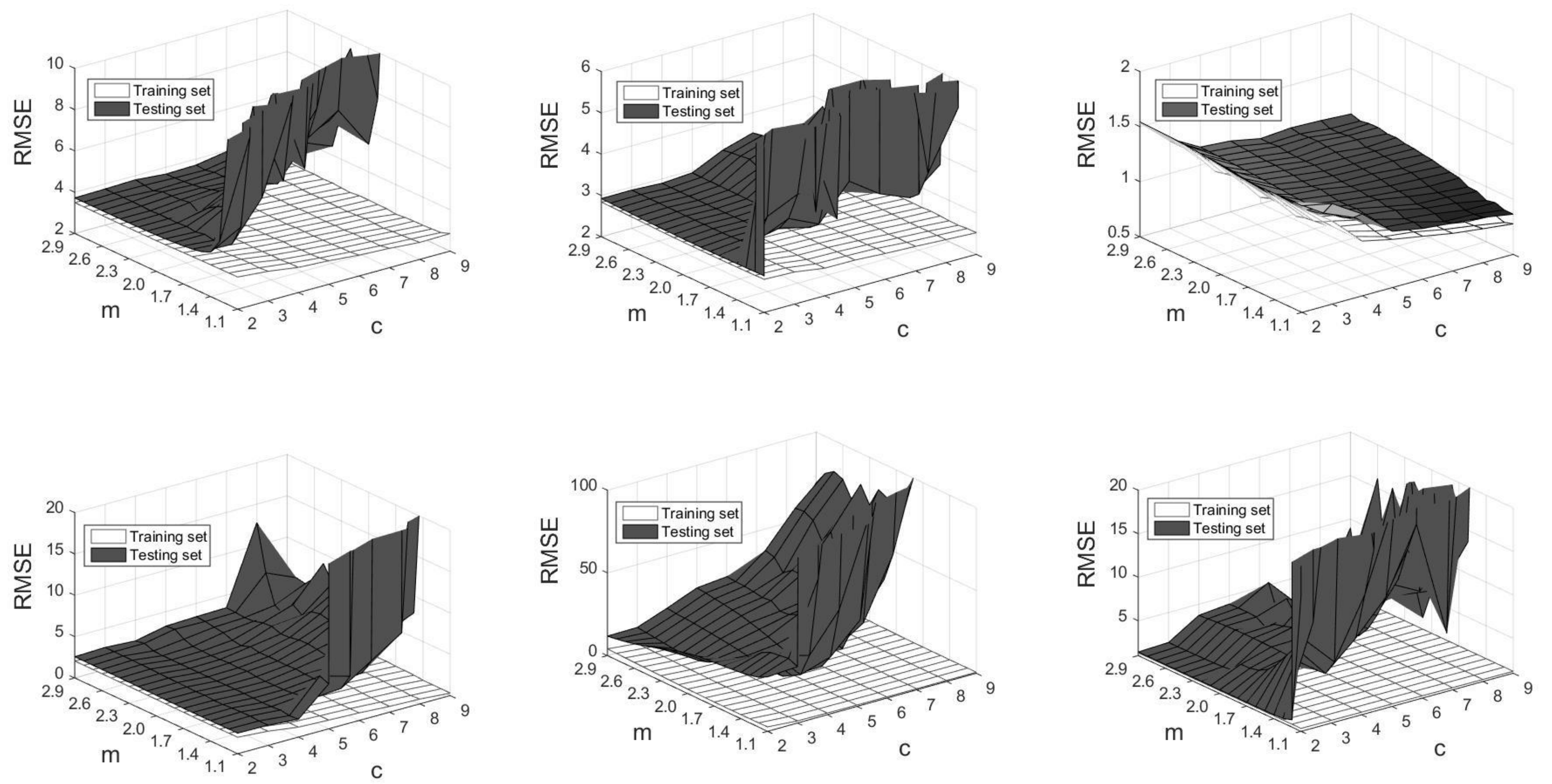
Name of the data	Number of instances	Number of inputs	Origin of the data
Boston housing	506	13	UCI machine learning repository https://archive.ics.uci.edu/ml/
Auto–MPG	392	7	UCI machine learning repository http://archive.ics.uci.edu/ml/
Stock	950	9	StatLib repository http://www.dcc.fc.up.pt/~ltorgo/Regression/
Concrete Slump	103	9	UCI machine learning repository http://archive.ics.uci.edu/ml/
Yacht Hydrodynamics	308	6	UCI machine learning repository http://archive.ics.uci.edu/ml/
Forest fires	517	12	UCI machine learning repository https://archive.ics.uci.edu/ml/

Experiments- Machine Learning Fuzzy model

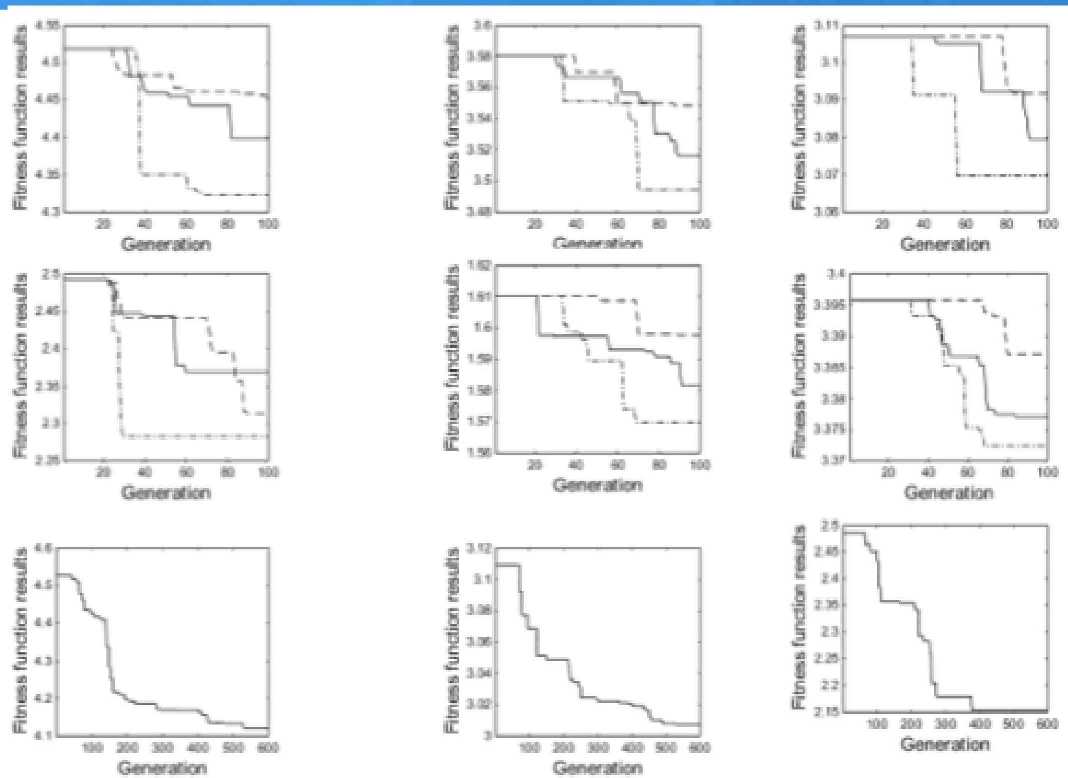
RMSE values of fuzzy models (c = 2, 5, 9 and m = 2.0)

		$c = 2$	$c = 5$	$c = 9$
Boston housing	Training	3.4892±0.0676	3.0078±0.0703	2.6388±0.0675
	Testing	3.7605±0.5430	5.0913±4.3772	5.3335±3.0400
Auto MPG	Training	2.7918±0.0540	2.6695±0.0496	2.5374±0.0614
	Testing	2.8698±0.4809	2.9212±0.4764	4.1705±2.6440
Stock	Training	1.4843±0.0134	1.0080±0.0234	0.9082±0.0104
	Testing	1.4905±0.1256	1.0766±0.0945	1.0144±0.0911
Concrete Slump	Training	1.9926±0.0652	1.2082±0.0857	0.5901±0.1236
	Testing	2.4860±0.6233	2.3828±0.7820	7.5163±3.6688
Yacht Hydrodynamics	Training	2.6702±0.0636	0.9445±0.0513	0.6528±0.0789
	Testing	24.2685±2.2786	14.7590±3.9352	76.5720±22.2185
Forest fires	Training	1.3312±0.0149	1.2883±0.0151	1.2315±0.0161
	Testing	1.4235±0.1696	4.1460±8.0625	6.1332±13.2331

Experiments- Machine Learning data sets

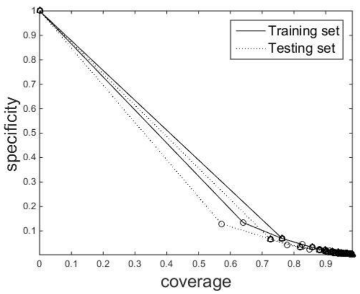


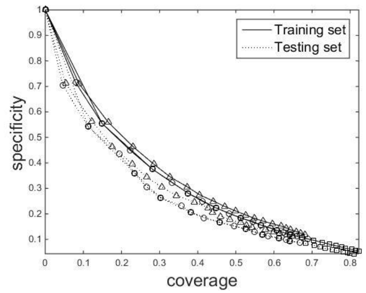
Experiments- Machine Learning data sets

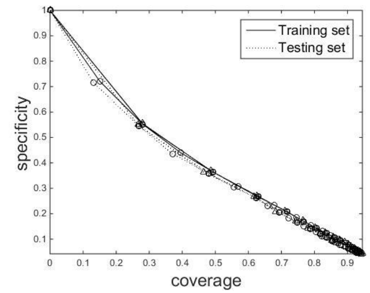


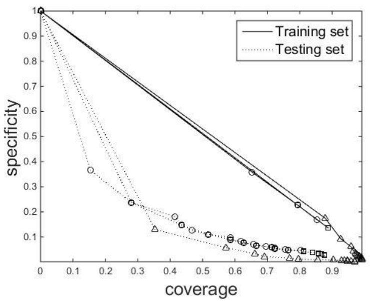
Performance of PSO for the third scenario of allocation of information granularity: (a) Boston housing (*c* = 2), (b) Auto MPG (*c* = 9), (c) Stock (*c* = 2), (d) Concrete Slump (*c* = 5), (e) Yacht Hydrodynamics (*c* = 9), (f) Forest fires (*c* = 5), (g) Boston housing (*c* = 2), (h) Stock (*c* = 2),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

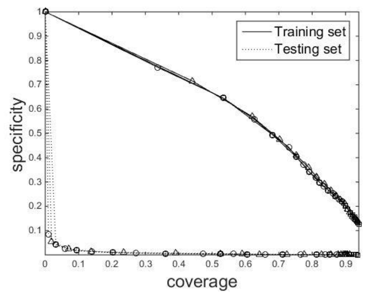
Experiments- Machine Learning data sets

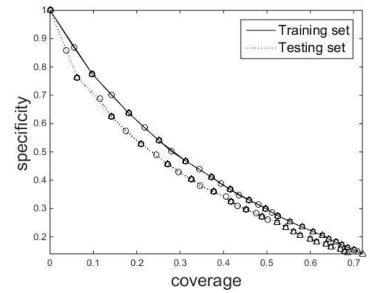




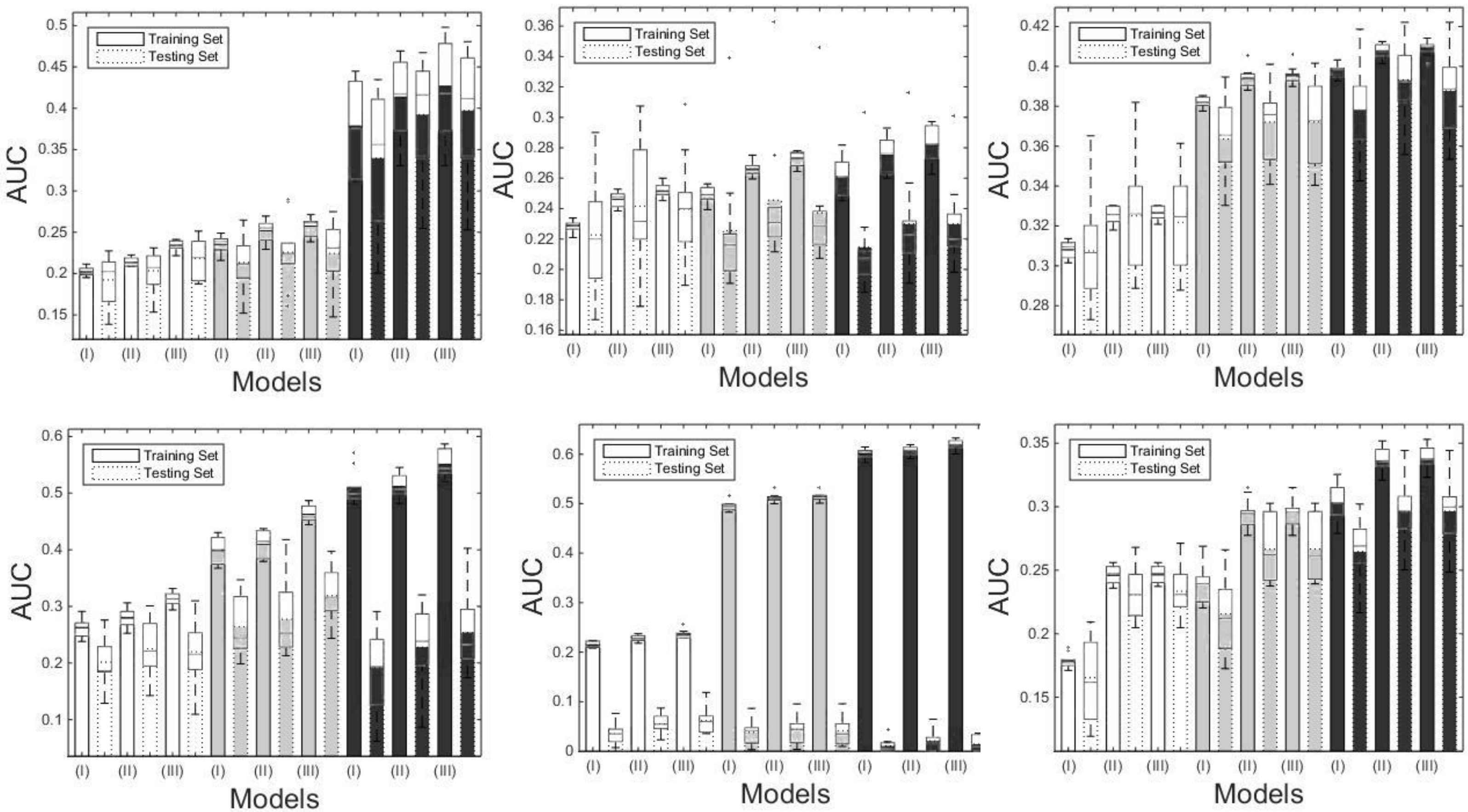






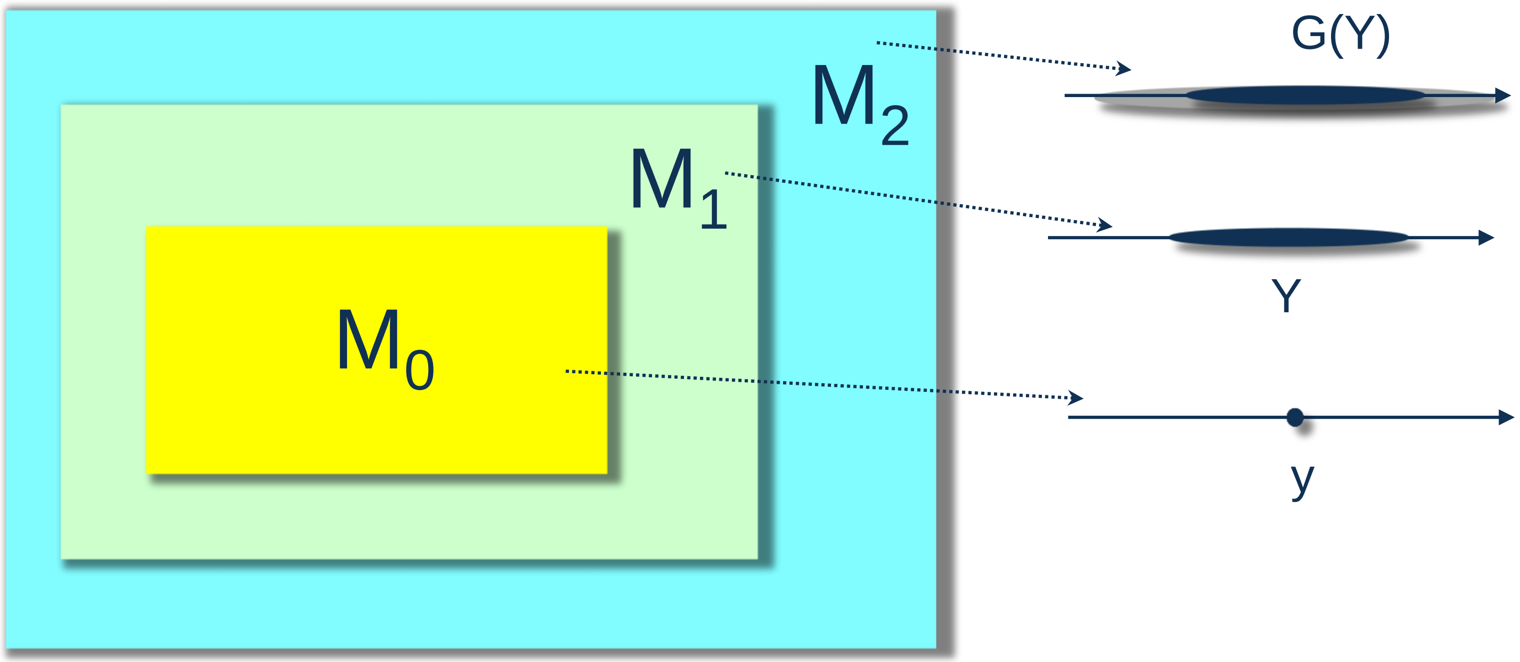


Experiments- Machine Learning data sets

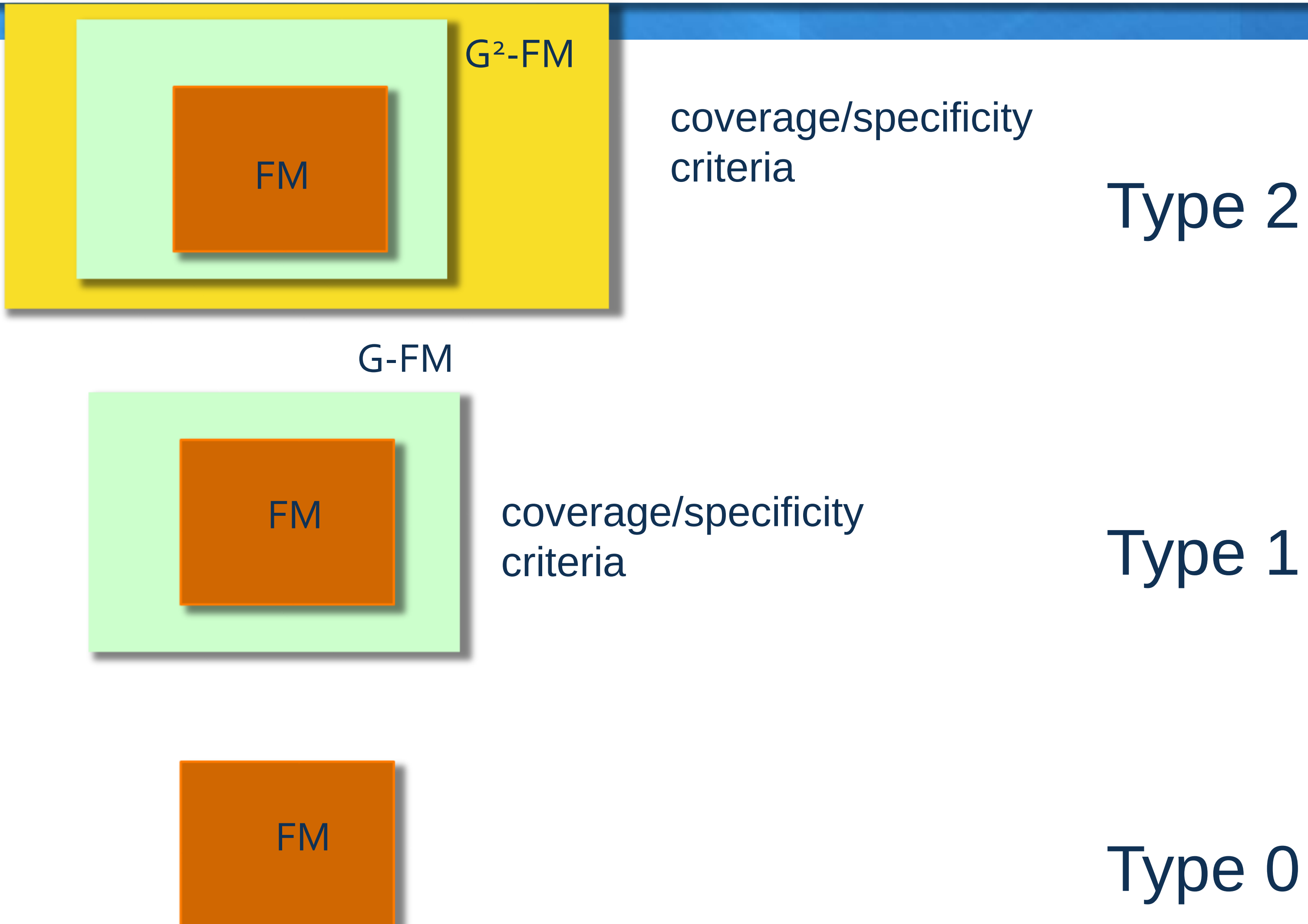


transparent bars -*c* = 2, gray bars - *c* = 5, and black bars – *c* = 9.

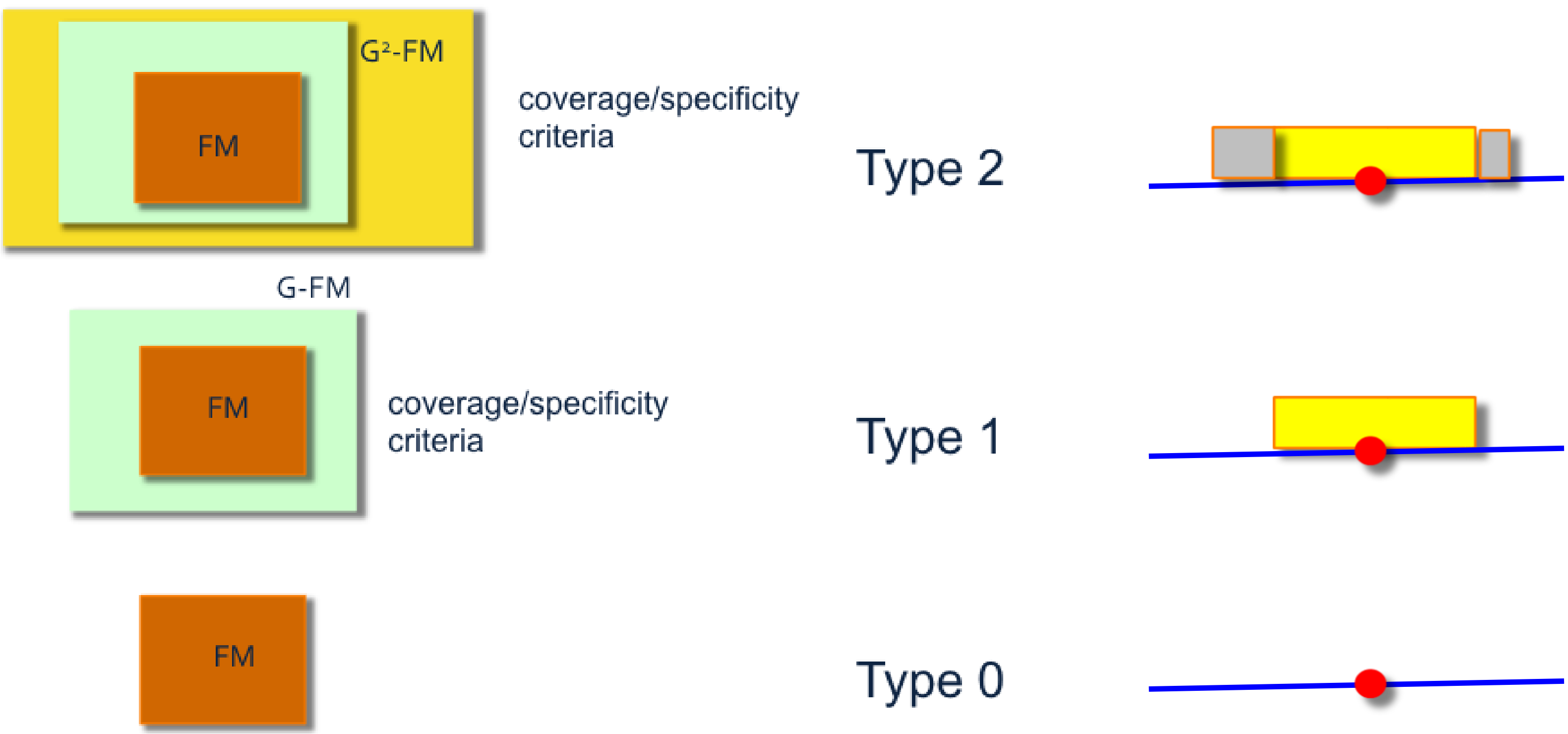
Hierarchy of granular models



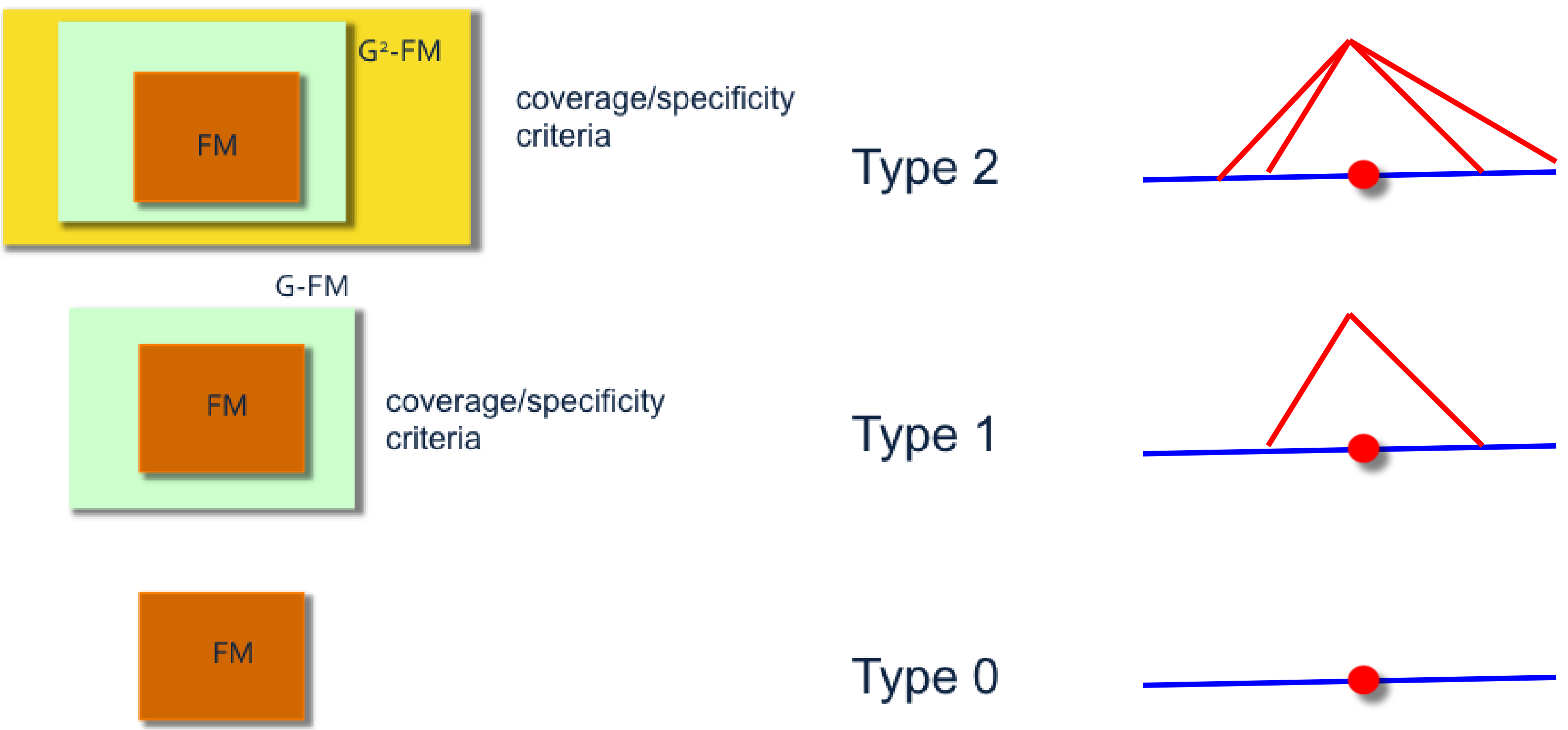
Granular fuzzy models of higher type



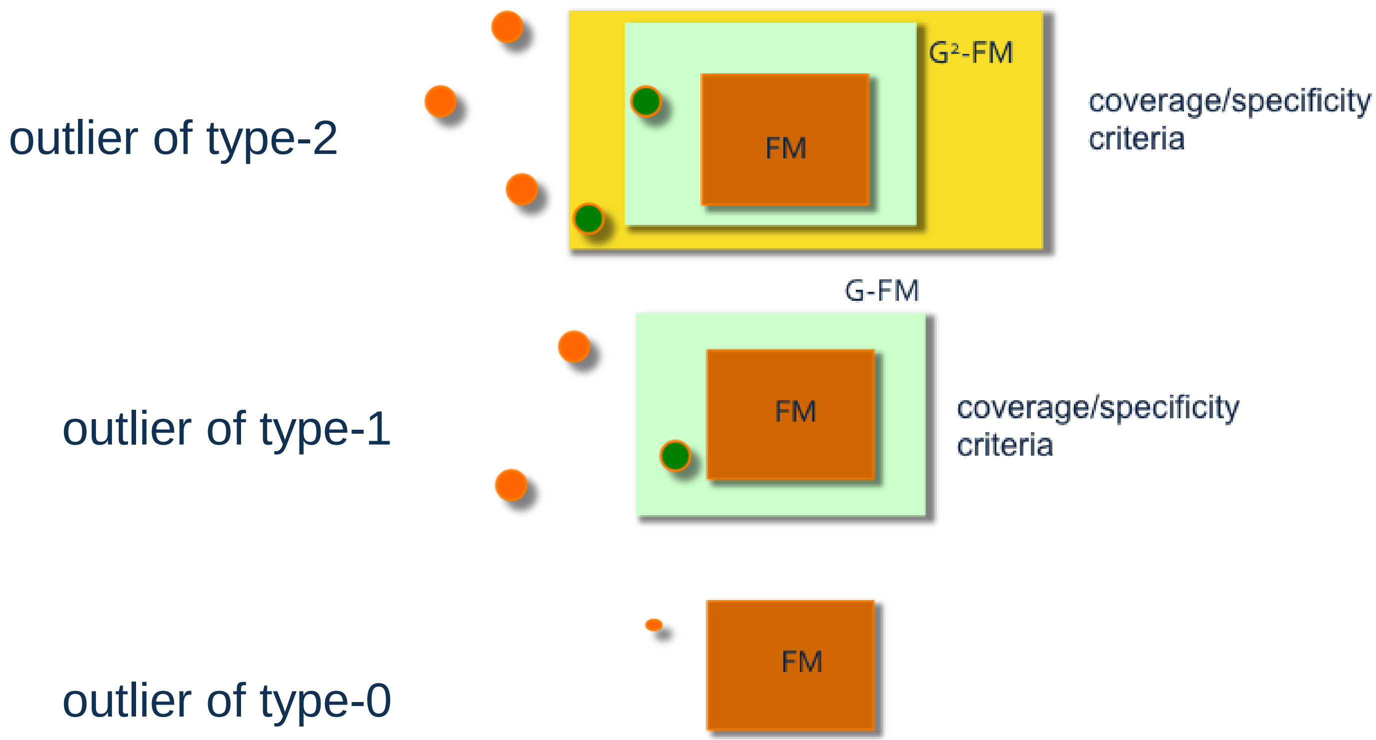
Granular fuzzy models granular intervals



Granular fuzzy models
Interval-valued fuzzy sets



Granular fuzzy models
Granular outliers



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…)