KNOWLEDGE-BASED CLUSTERING

Witold PEDRYCZ

Department of Electrical & Computer Eng.
University of Alberta,Edmonton, Canada
&
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

March 2005

Agenda

Introduction: Granular Information Processing
Formal frameworks of Granular Computing
Knowledge-based clustering and information granulation
Facets of knowledge-based clustering
- Partial supervision
- Proximity-based clustering
- Implicit knowledge guidance

Context-oriented guidance
Collaborative clustering
Linguistic (granular) models
Conclusions

Information granules and image processing

Images → perception and understanding

Meaningful entities (objects)

GRANULATION OF SPATIAL INFORMATION

Numeric information (pixels)

Images: from processing to understanding

User, Decision-maker, Designer,…

Symbols

UNDERSTANDING

RELEVANCE FEEDBACK

SEMANTIC GAP

Numbers

Numeric information (pixels) and processing

From images to their interpretation

Understanding — **INTERPRETATION RECOGNITION CONTENT-BASED RETRIEVAL**

Reconciliation -bridging numbers and symbols

Granular constructs — **GRANULAR COMPUTING**

Images — **IMAGE PROCESSING (segmentation, filtering,edge detection…)**

Time series - examples

$x(t+1) = f(x(t), x(t-1))$ $x(t+1) = a_0 + a_1 x(t-1) + a_2 x(t)$

LINEAR OR NONLINEAR MODELS
NUMERIC IN TIME AND SPACE

Goal-oriented models of time series

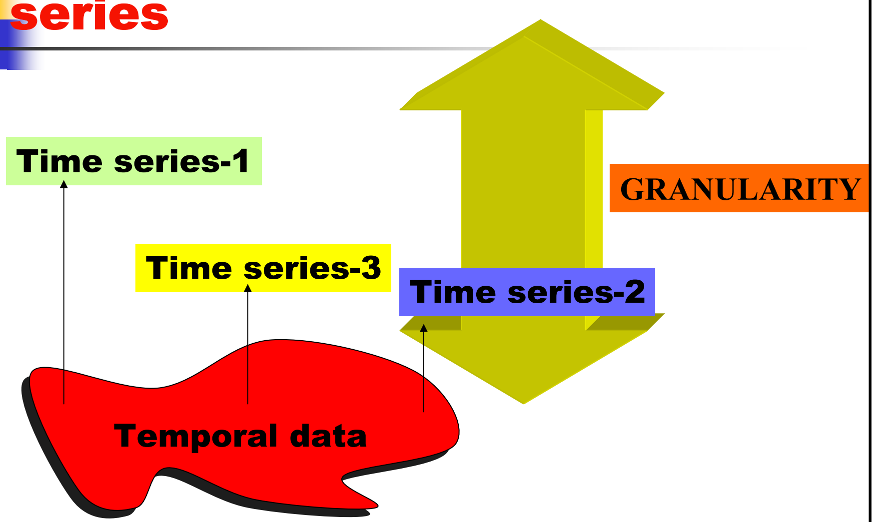
Visualize <u>dominant</u> temporal relationships

Provide a qualitative description of <u>interesting</u> dependencies

Compare several time series- are they qualitatively <u>similar?</u>

Context (task) dependent
User driven
Goal and user-centric models

Granular(fuzzy set-based) time series



Granular Computing

Granular Computing

Information granules as semantic entities

Information granulation

Formal frameworks

Information granules

Semantically meaningful entities composed of elements drawn together on a basis of

- similarity,
- functional closeness,
- spatial neighborhood, etc.

and regarded as generic elements in any processing pursuits (viz. granular computing)

Information granulation

Process of forming information granules using domain knowledge and experimental evidence and exploiting one of formal mathematical frameworks (logic, set theory, fuzzy sets, probability…)

Development of Information Granules

Information granules: Goal

Granulate data that is construct

<u>semantically meaningful</u> entities

supporting <u>efficient design</u> of models

Numeric data and information granules

Numeric data

(a) abundance in time and space
(b) difficult to interpret
(c) no explicit semantics

[agenda of data mining]

Development approaches

- Designer-centric
- Data centric
- Hybrid schemes

Development approaches Designer-centric

Information granules specified by designer/user at the level of:

- general type of membership functions or membership grades at individual points
- parameters of membership functions (modal values, spreads)
- number of information granules (7+/-2)
- data not (explicitly) involved in the constructs

PRESCRIPTIVE DESIGN APPROACH

Development approaches Data-centric

Information granules specified on a basis of experimental numeric data

- Predominantly related to clustering (being regarded as a basic algorithmic vehicle of forming fuzzy sets or fuzzy relations), see Fuzzy C-Means (FCM)
- Reflects structure in data; are these clusters semantically meaningful?

DESCRIPTIVE DESIGN APPROACH

Development approaches Hybrid (knowledge-based)

Information granules are specified on a basis of experimental numeric data and reflect the semantics conveyed by the designer

Clustering and Knowledge

Information granules and clustering

Data → clustering → information granules (clusters)

Clustering and Knowledge

Data centric approach → clustering

Human centric approach → knowledge hints

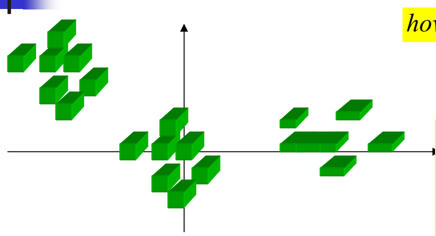
→ Knowledge-based clustering

Fuzzy C-Means as Algorithmic Framework of Clustering

- Well-developed optimization environment – typical objective function clustering approach
- Commonly used
- Lot of comparative studies

Structure representation

how to represent clusters(groups)?



Partition matrix

N patterns

c groups

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Partition matrix

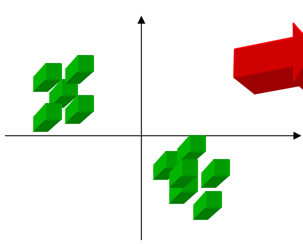
$$U = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

cluster - 1: $\{1,4,5,8\}$

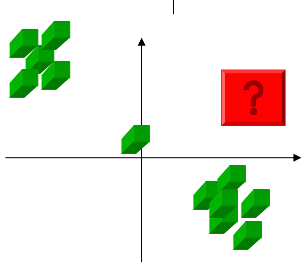
cluster - 2 : $\{2,3\}$

cluster - 3 : $\{6,7\}$

Partition matrix



$$u_{ik} = \begin{cases} 1 & \text{if } x_k \text{ is in i - th cluster} \\ 0, & \text{otherwise} \end{cases}$$



$u_{ik} \in [0,1]$

Partition matrix

U satisfies the following conditions:

$$- \sum_{i=1}^{c} u_{ik} = 1 \ \text{ for all } k = 1, 2, ..., N$$

$$- 0 < \sum_{k=1}^{N} u_{ik} < N \ \text{ for all } i = 1, 2, ..., c$$

Family of partition matrices $\mathcal{U}$

Objective function

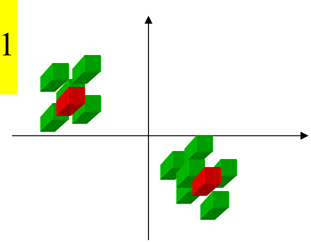
FCM (Fuzzy C-Means, Bezdek,1981)

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \| \mathbf{x}_k - \mathbf{v}_i \|^2, \quad m > 1$$

$\mathbf{v}_i$: prototypes

$U = [u_{ik}]$: partition matrix

m - fuzzification coefficient



$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \| \mathbf{x}_k - \mathbf{v}_i \|^2 \Rightarrow \mathrm{Min}_{\text{prototypes}, U \in \mathbf{U}} Q$$

FCM – detailed calculations

Use of techniques of Lagrange multipliers to accommodate constraints (partition matrix)

$$V = \sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2} - \lambda\left(\sum_{i=1}^{c} u_{ik} - 1\right)$$

$$\frac{\partial V}{\partial u_{st}} = 0, \quad \frac{\partial V}{\partial \lambda} = 0 \quad s = 1,2, \ldots, c, \; t = 1, 2, \ldots, N$$

FCM – flow of optimization

repeat

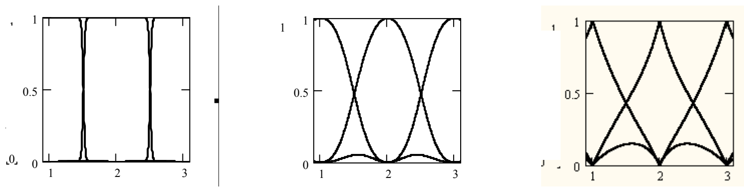
$$u_{st} = \frac{1}{\sum_{i=1}^{c} \left(\frac{d_{st}}{d_{it}} \right)^{\frac{2}{m-1}}}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^{m} \mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^{m}}$$

$$d_{st} = \| \mathbf{x}_t - \mathbf{v}_s \|$$

Until stopping criterion satisfied

Fuzzification parameter in FCM



m =1.1 m=2.0 m=3.0

m=2.0 commonly used value

Mechanisms of knowledge-based clustering

Partial supervision

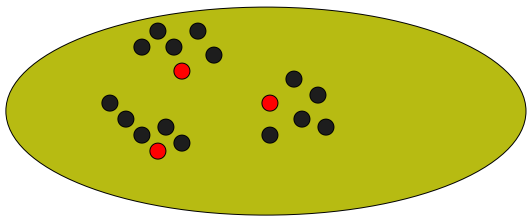
Proximity clustering

Context-based clustering

Collaborative clustering

Partial supervision in Fuzzy C-Means (FCM)

- High number of data
- Some patterns are labeled
- How to use this information (supervision hints) in the clustering procedure- navigation of FCM



Partially supervised FCM

Number of clusters = number of classes

Augmented objective function

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 \| \mathbf{x}_k - \mathbf{v}_i \|^2 + \alpha \underbrace{\sum_{i=1}^{c} \sum_{k=1}^{N} (u_{ik} - b_k f_{ik})^2 \| \mathbf{x}_k - \mathbf{v}_i \|^2}_{\text{Supervision- coming from labeled patterns}}$$

$\alpha > 0$: impact of the supervision part

Partially supervised FCM

$$Q=\sum_{i=1}^{c}\sum_{k=1}^{N}u_{ik}^{2}\,\|\mathbf{x}_k-\mathbf{v}_i\|^2+\alpha\underbrace{\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik}-b_k f_{ik})^2\,\|\mathbf{x}_k-\mathbf{v}_i\|^2}$$

$$b_k=\begin{cases}1 & \text{if k-th pattern is labeled}\\ 0 & \text{otherwise}\end{cases}$$

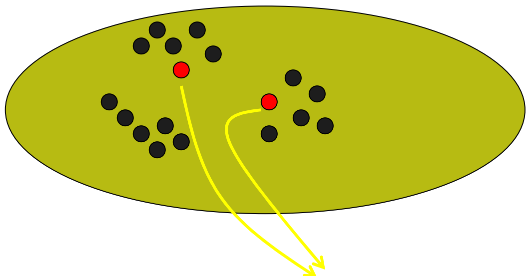
f_{ik} – class assignment (label)

Partial supervision: Optimization process

$$Q=\sum_{i=1}^{c}\sum_{k=1}^{N}u_{ik}^{2}\,\|\mathbf{x}_k-\mathbf{v}_i\|^2+\alpha\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik}-b_k f_{ik})^2\,\|\mathbf{x}_k-\mathbf{v}_i\|^2$$

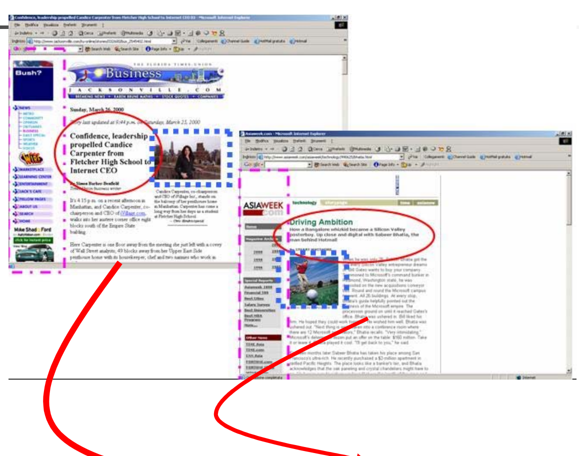
$$\begin{cases} u_{ik}=\dfrac{1}{1+\alpha}\left[\dfrac{1+\alpha\left(1-b_k\sum_{i=1}^{c}f_{ik}\right)}{\sum_{j=1}^{c}\left(\dfrac{d_{ik}}{d_{jk}}\right)^2}+\alpha f_{ik}b_k\right] \\[2ex] \mathbf{v}_s=\dfrac{\sum_{k=1}^{N}\psi_{sk}\mathbf{x}_k}{\sum_{k=1}^{N}\psi_{sk}} \qquad \psi_{ik}=[u_{ik}^2+(u_{ik}-f_{ik}b_k)^2 \end{cases}$$

Proximity – based FCM



proximity of patterns = λ

Proximity – based FCM



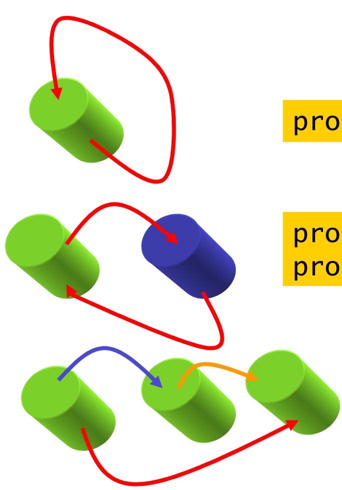
proximity of patterns = λ

Proximity – based FCM: motivation

"unknown" distance function

incomplete feature space (multimedia)

Proximity measure



proximity (pattern-i, pattern-i)= 1

proximity (pattern-i, pattern-j)=
proximity (pattern-j, pattern-i)

transitivity not required

Partition matrix and proximity measure

Patterns "k_1" and "k_2" with membership grades included in partition matrix U

$$\hat{p}[k_1,k_2] = \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2})$$

$$\Rightarrow \quad \hat{p}[k_1,k_1] = \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_1}) = \sum_{i=1}^{c} u_{ik_1} = 1$$

$$\Rightarrow \quad \hat{p}[k_1,k_2] = \hat{p}[k_2,k_1]$$

Proximity – based FCM

Combined objective function

$$\left\{ \begin{array}{l} Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 \|\mathbf{x}_k - \mathbf{v}_i\|^2 \\ V = \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} (\hat{p}[k_1,k_2] - p[k_1,k_2])^2 b[k_1,k_2] \|\mathbf{x}_{k_1} - \mathbf{x}_{k_2}\| \end{array} \right.$$

$$\hat{p}[k_1,k_2] = \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2})$$

proximity

Proximity – based FCM: Optimization processes

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^2 \| \mathbf{x}_k - \mathbf{v}_i \|^2$$

$$V = \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} (\hat{p}[k_1,k_2] - p[k_1,k_2])^2 b[k_1,k_2] \| \mathbf{x}_{k_1} - \mathbf{x}_{k_2} \|$$

Standard optimization flow of FCM

Gradient-based optimization w.r.t. U

$$U(iter+1) = U(iter) - \gamma \nabla_U V$$

Proximity – based FCM: Gradient-based optimization

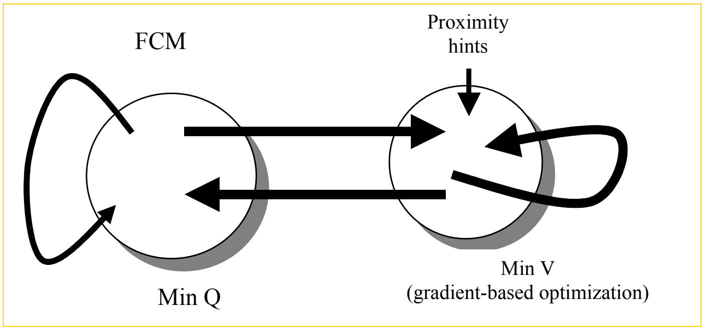
$$\frac{\partial V}{\partial u_{st}(iter)} = \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \frac{\partial}{\partial u_{st}} (\sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2}) - p[k_1,k_2])^2 b[k_1,k_2] d[k_1,k_2] =$$

$$= 2 \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2}) - p[k_1,k_2]) b[k_1,k_2] d[k_1,k_2] \frac{\partial}{\partial u_{st}} \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2})$$

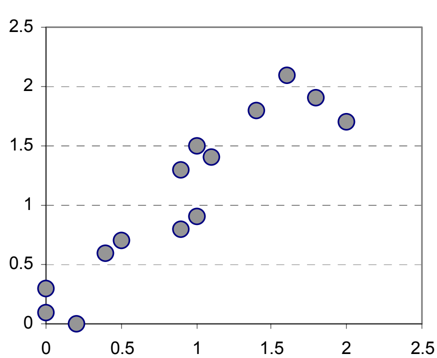
$$\varphi[s,t,k_1,k_2] = \frac{\partial}{\partial u_{st}} \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2}) = \begin{cases} 1 \text{ if } t = k_1 \text{ and } u_{sk_1} \le u_{sk_2} \\ 1 \text{ if } t = k_2 \text{ and } u_{sk_2} \le u_{sk_1} \\ 0 \qquad \text{otherwise} \end{cases}$$

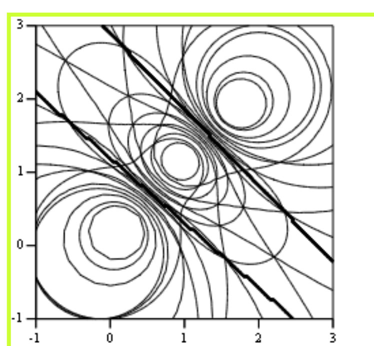
$$\frac{\partial V}{\partial u_{st}(iter)} = 2 \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \sum_{i=1}^{c} (u_{ik_1} \wedge u_{ik_2}) - p[k_1,k_2]) \quad \varphi[s,t,k_1,k_2]$$

Proximity clustering : Overall Optimization

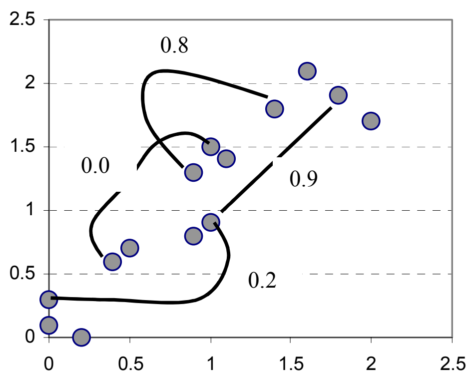


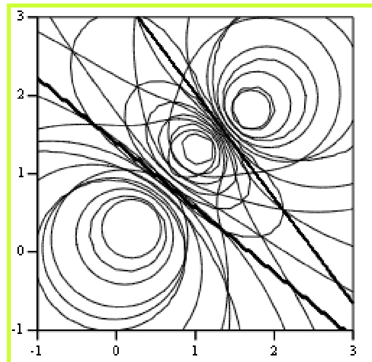
Example – Synthetic data (1)





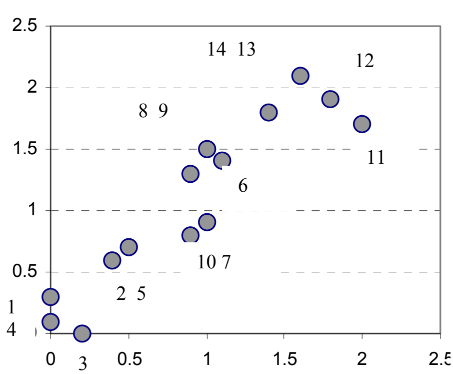
Example – Synthetic data (2)

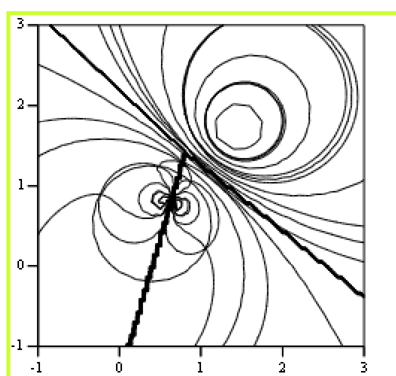




Example – Synthetic data (3)

Proximity hints: (1 7 0.9), (2 6 0), (2 9 1), (7 12 0.9), (9 13 0.0), (8 14 0), (13 14 0), (1 2 0)





Implicit knowledge guidance: a general perspective

- **Proximity** hints :Prox(x, y)
  - Prox(x,y) *equal* to λ
  - Prox(x,y) *less than* (*greater than*) μ
- **Uncertainty** [entropy] hints: H(x)
  - H(x) *equal* to λ
  - H(x) *less than* (*greater than*) μ
- ....

Conditional (context-based) clustering

<u>Generic task</u>

Cluster data

<u>Specialized task (context)</u>

Cluster data in context A

Domain(problem-oriented) knowledge

Conditional (context-based) clustering

Structure in a database of customers
given *high income customers*

$\underbrace{\hspace{6em}}_{\text{Context A}}$

Structure in a database of customers
given all customers

$\underbrace{\hspace{6em}}_{\text{no context}}$

Conditional (context-based) clustering

Objective function Q

Context-based partition matrix

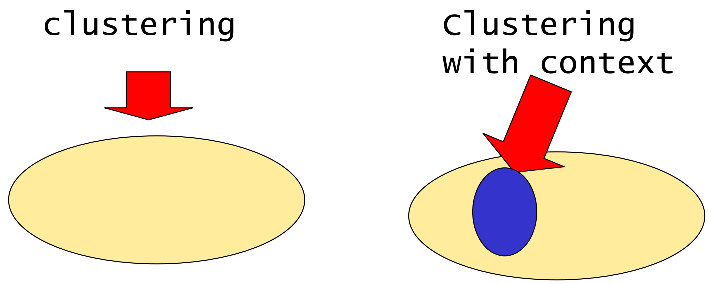
$$\sum_{i=1}^{c} u_{ik} = A(y_k)$$

Context (A) at y_k

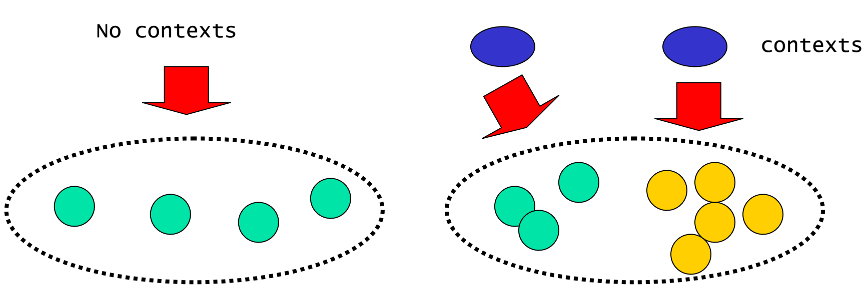
Conditional (context-based) clustering

Context – as semantic filter

Focus on some fuzzy subset of data (implied by specified context)



Conditional (context-based) clustering



Partition matrix with context

U satisfies the following conditions:

$$- \sum_{i=1}^{c} u_{ik} = A(y_k) \text{ for all } k = 1, 2, \ldots, N$$

$$- 0 < \sum_{k=1}^{N} u_{ik} < N \text{ for all } i = 1, 2, \ldots, c$$

Family of partition matrices *U(A)*

FCM – flow of optimization

repeat

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^m}$$

$$u_{st} = \frac{A(y_t)}{\sum_{i=1}^{c} \left(\frac{d_{st}}{d_{it}} \right)^{2/(m-1)}}$$

$$d_{st} = \| \mathbf{x}_t - \mathbf{v}_s \|$$

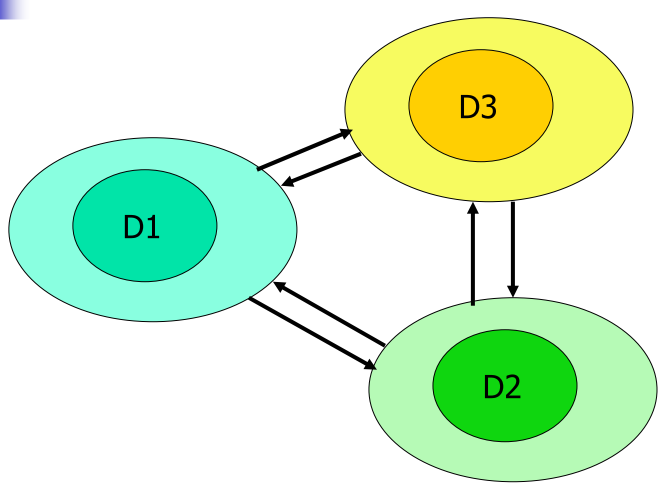
Until stopping criterion satisfied

Collaborative clustering

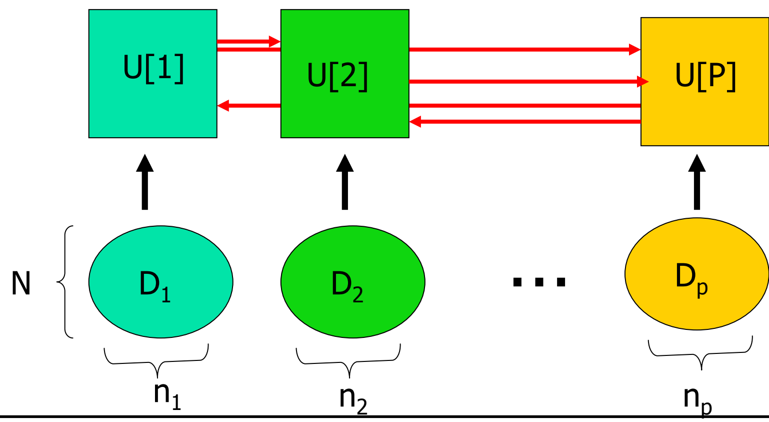
Design of information granules on a basis of several disjoint sources of data

Collaborative effort under limited level of sharing of data (communication realized at some level of information granularity)

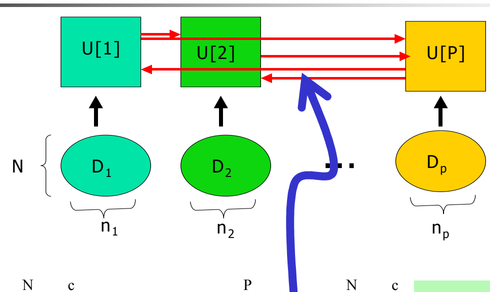
Collaborative clustering



Collaborative clustering horizontal mode



Collaborative clustering: horizontal mode



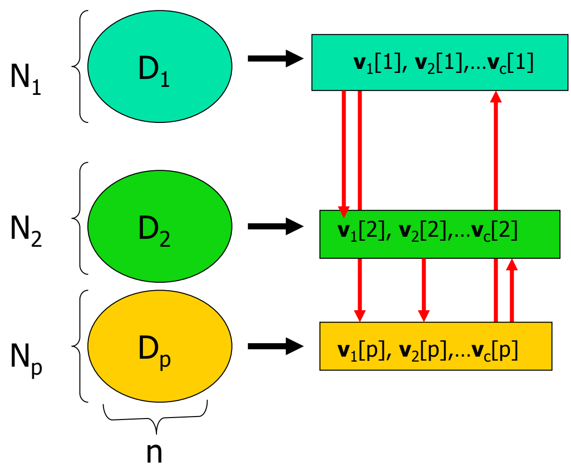
$$Q[ii] = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^2[ii] d_{ik}^2[ii] + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii,jj] \sum_{k=1}^{N} \sum_{i=1}^{c} \{u_{ik}[ii] - u_{ik}[jj]\}^2 d_{ik}^2[ii]$$

$$u_{st}[ii] = \frac{\varphi_{st}[ii]}{1+\psi[ii]} + \frac{1}{\sum_{j=1}^{c} \frac{d_{st}^2}{d_{jt}^2}} [1 - \sum_{j=1}^{c} \frac{\varphi_{jt}[ii]}{1+\psi[ii]}]$$

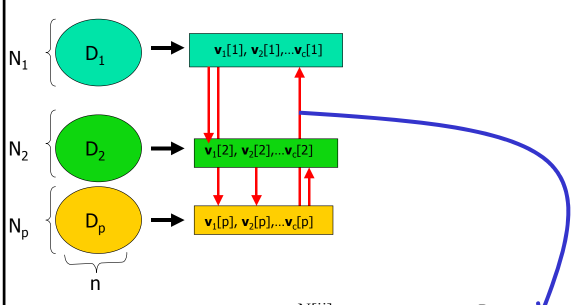
$$\varphi_{st}[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii,jj] u_{st}[jj]$$

$$\psi[ii] = \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \alpha[ii,jj]$$

Collaborative clustering: vertical mode



Collaborative clustering: vertical mode



$$Q[ii] = \sum_{i=1}^{c} \sum_{i=1}^{N[ii]} u_{ik}^2[ii] d_{ik}^2 + \sum_{\substack{jj=1 \\ jj \neq ii}}^{P} \beta[ii.jj] \sum_{k=1}^{N[ii]} \sum_{i=1}^{c} u_{ik}^2[ii] \| v_i[ii] - v_i[jj] \|^2$$

Linguistic(granular) models

Design objective

Construct a model at the level of information granules

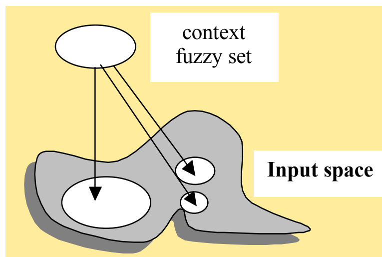
User-centric constructs – the designer plays an active role in model development

Results produced by model directly reflective of the level of granularity of the model

Overall design

Data – a view from a certain user-defined perspective:

Focal information granules → blueprint of model



Context-based clustering

Given contexts (fuzzy sets) in output space

Cluster data in input space

Web of information granules (context-induced fuzzy sets)

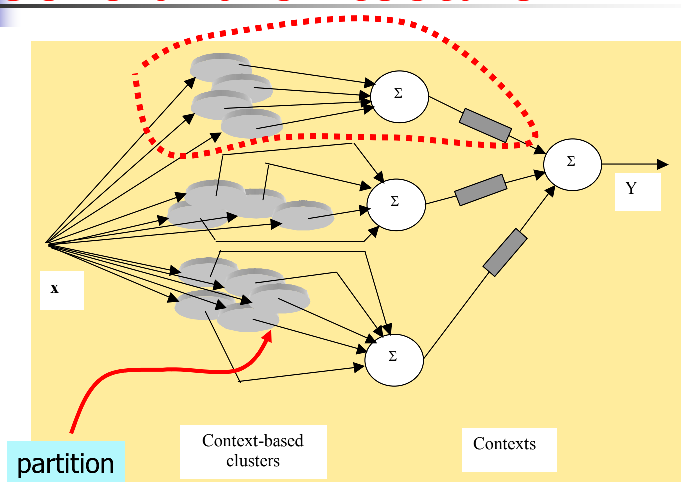
Linguistic models

Web of information granules
(context-induced fuzzy sets)

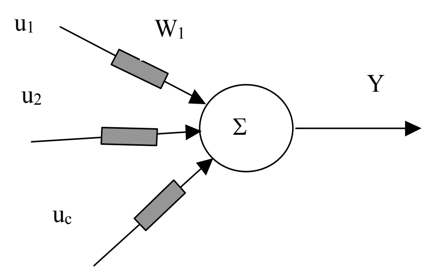
Forming dependencies between information granules

Linguistic model:web of connections(links)

General architecture



Granular neuron (1)



$$Y = N(u_1, u_2, .., u_c, W_1, W_2, .., W_c) = \sum_{\oplus} W_i \otimes u_i$$

Granular neuron (2)

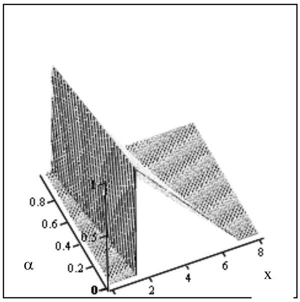
two inputs: u1 $=\alpha$, u2= $1-\alpha$

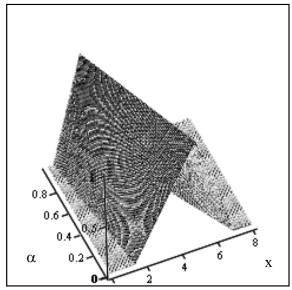
Triangular connections: W1 and W2

Output of neuron

$$\sum_{i=1}^{c} a_i u_i, \sum_{i=1}^{c} m_i u_i, \sum_{i=1}^{c} b_i u_i$$

Granular neuron (3)





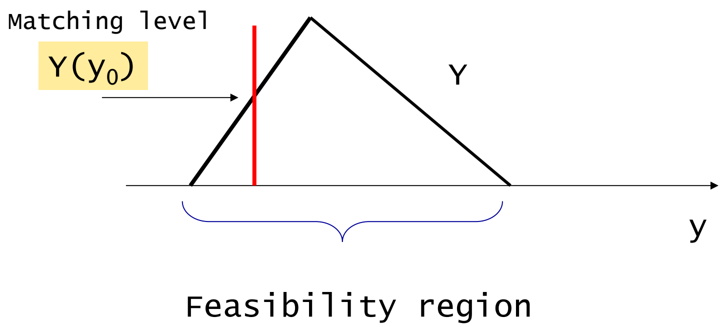
Linguistic network: interpretation

Output (Y) as information granule

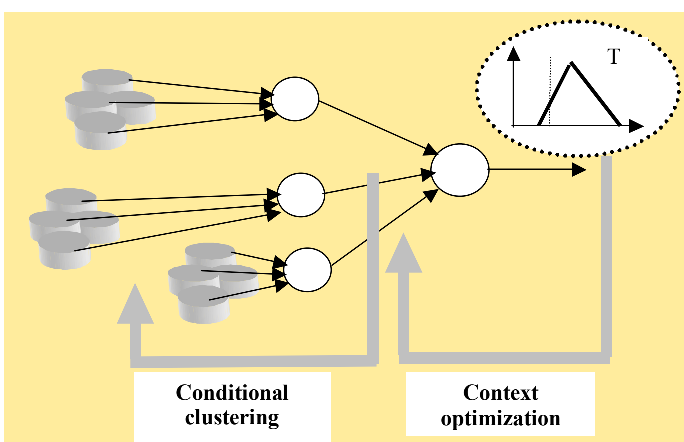
- Visualization of possible outcomes of the model with membership degrees

- Matching model and experimental data

Linguistic network: interpretation

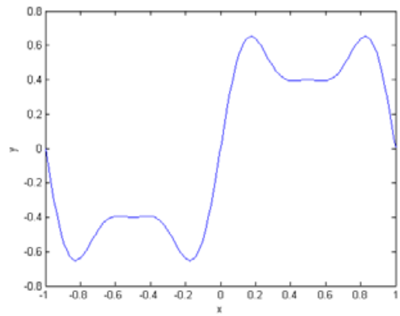


Refinement of the model

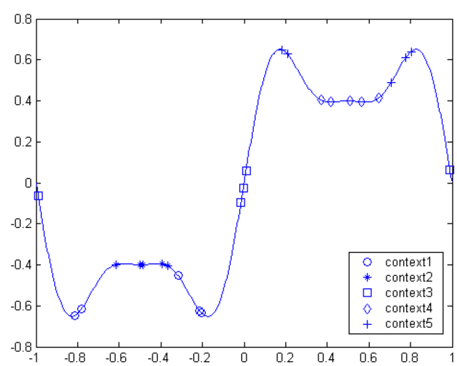


Numeric experiment

$$y = 0.6\sin(\pi x) + 0.3\sin(3\pi x) + 0.1\sin(5\pi x)$$

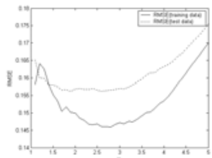


Numeric experiment

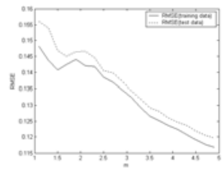


Parametric studies: fuzzification factor

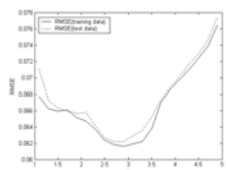
No of contexts =3
No of clusters =10



No of contexts =10
No of clusters =2



No of contexts =5
No of clusters =5



Conclusions

Information granules as generic constructs supporting functionality of human-centric systems

Knowledge-based clustering as a conceptual and algorithmic environment for information granulation

Diversity of mechanisms of knowledge-based enhancements of information granulation