Data Projection and Visualization Using Self-Organizing Neural Networks

Pablo A. Estévez

Depto. Ingeniería Eléctrica

Universidad de Chile

EVIC 2005

Contents

- Vector Quantization
- Self-Organizing Neural Networks
  - Kohonen Maps
  - Neural Gas
- Data projection
  - Sammon's mapping
- Online Visualization
- Application to Information Retrieval
  - WEBSOM
- Conclusions

Unsupervised Learning

- The desired outputs (classes) are unknown
- Patterns could be grouped (clustered) by using similarity or proximity measures
- Example: Is there any relationship between a country Gross National Product, the level of people's illiteracy and children‘s mortality rate?
  - Each country is represented by a 3D feature vector

Vector Quantization

- The objective is to approximate a given training sample (or unknown generating distribution)

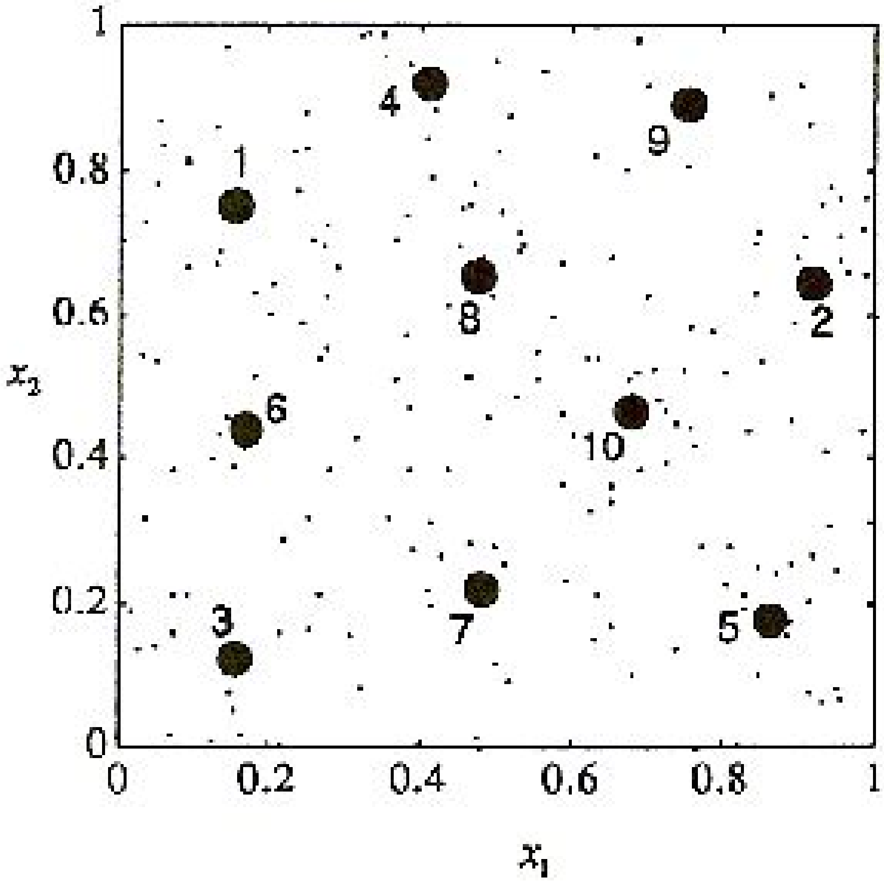
  where x_i is n-dim $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$
- Using a finite number of prototype (codebook) vectors

  where

$$\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_m\}$$

$$m < n$$

VQ Example in 2D



- Small points represent the data samples and large points indicate the prototype vectors

Applications of VQ

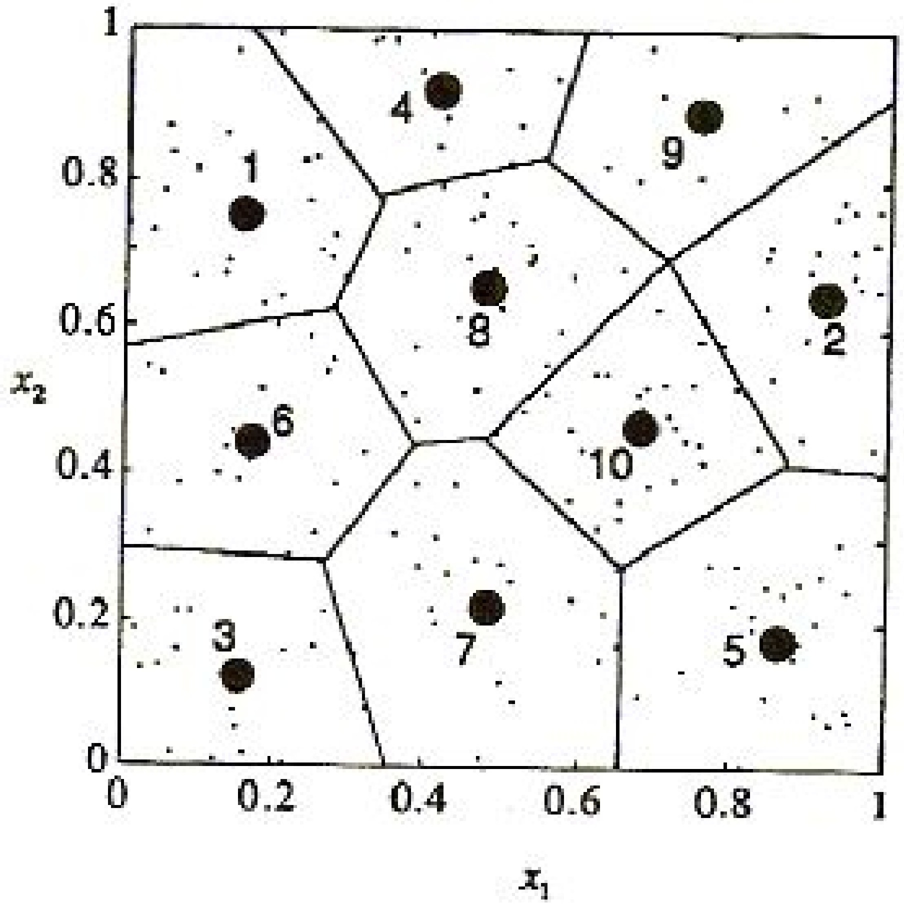
- Dimensionality reduction: to produce a compact/low-dimensional encoding of a given data set
- Interpretation of high-dimensional data sets
- Predictive modeling: to produce a model for future samples of the same distribution
- Preprocessing for supervised learning: feature extraction

Optimal VQ

- There are two necessary conditions for an optimal vector quantizer
  - The codebook vectors must be the centroids of each partition region
  - The partition mus $c_j = E\left[\mathbf{x} \mid \mathbf{x} \in R_j\right]$ or partition, also known as Voronoi regions

$$R_j = \left\{\mathbf{x} \in \Re^d : \left\|\mathbf{x} - \mathbf{c}_j\right\| < \left\|\mathbf{x} - \mathbf{c}_k\right\|\right\} \quad \forall k \neq j$$

Voronoi Partition



- The VQ partitions are non-overlapping and cover the entire input space

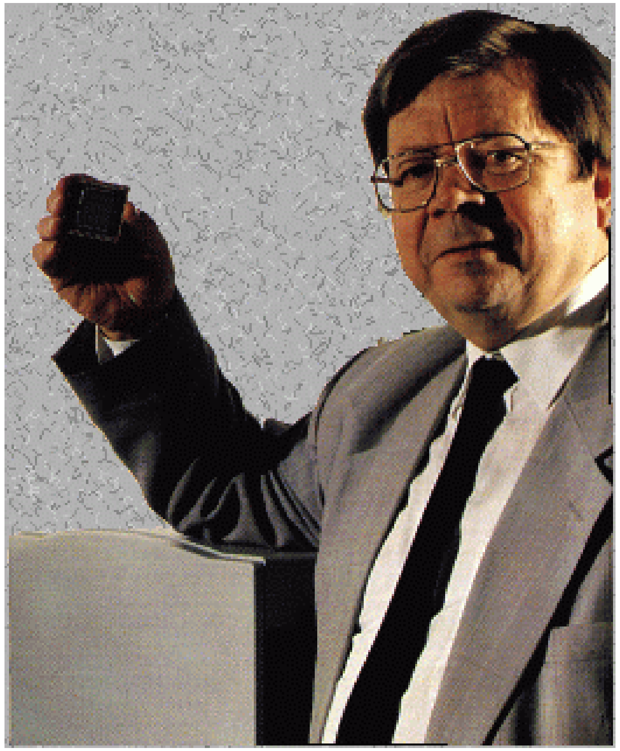
Clustering

- Clusters are continuous regions in feature space containing a relatively high-density of points, separated by regions of relatively low-density points
- The aim is to organize a collection of patterns into clusters based on a similarity measure
- Clustering NN use unsupervised learning

Kohonen's Map

- Self-Organizing Feature Map (SOM)
- Unsupervised Learning
- Vector Quantization of Feature Space
- Topological Ordered Mapping
- Main Applications:
  - Dimensionality reduction
  - Visualization of high-dimensional data in 2D or 3D maps
  - Clustering
  - Knowledge discovery in Databases

Teuvo Kohonen



http://www.cis.hut.fi/teuvo/

Kohonen's Map

- It defines a two-dimensional grid in output space
- Neighborhood is measured in the output space
- Each node in the grid has associated a vector prototype in input space

Kohonen's Map

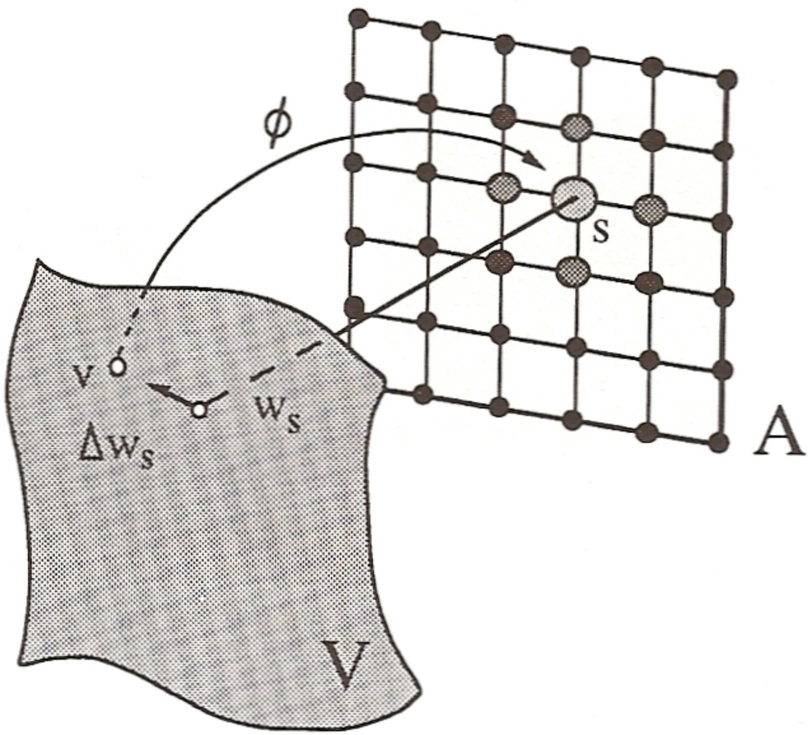


Figure 4.2 The adaptation step in Kohonen's model. The input value **v** selects a center **s** in whose neighborhood all neurons shift their weight vectors $\mathbf{w_s}$ towards the input **v**. The magnitude of the shift decreases as the distance of a unit from the center **s** increases. In the figure, this magnitude is indicated by different sizes and gray values. The shift of weights is only depicted, though, for unit **s**.

Topological Neighborhood

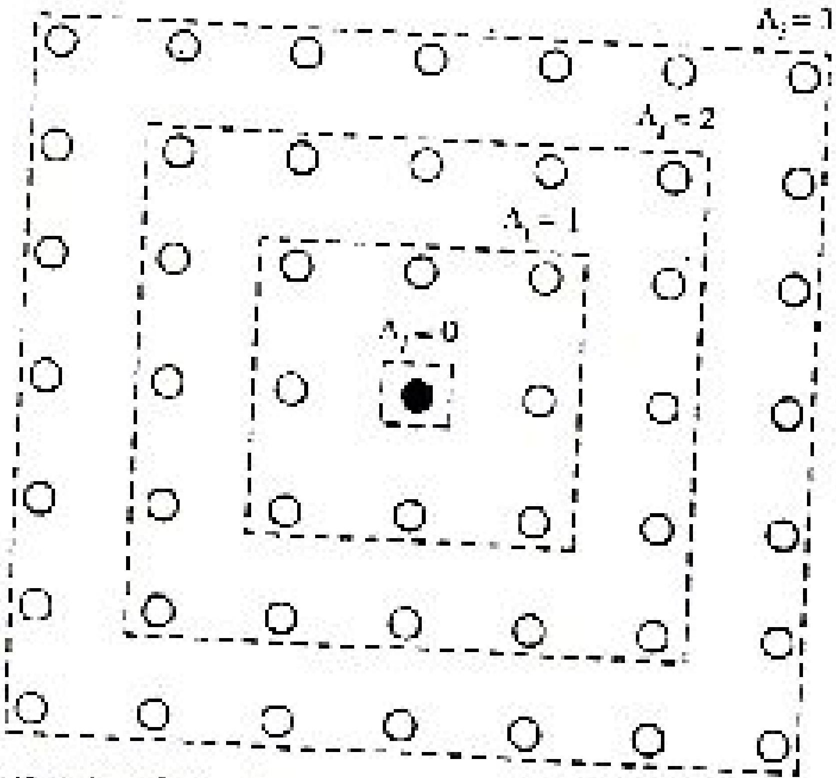
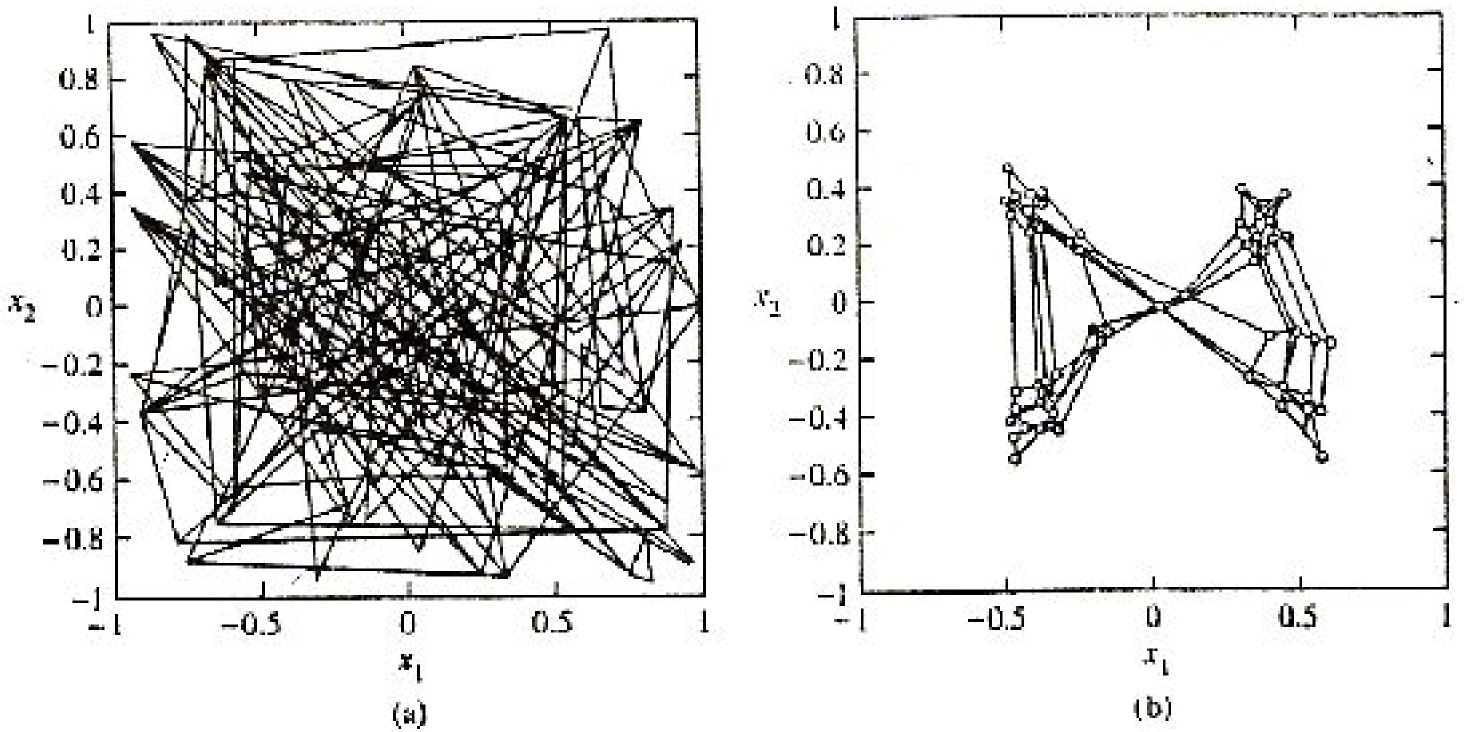


FIGURE 10.11 Square topological neighborhood Λ_i, of varying size, around "winning" neuron i, identified as a black circle.

Ordering Process



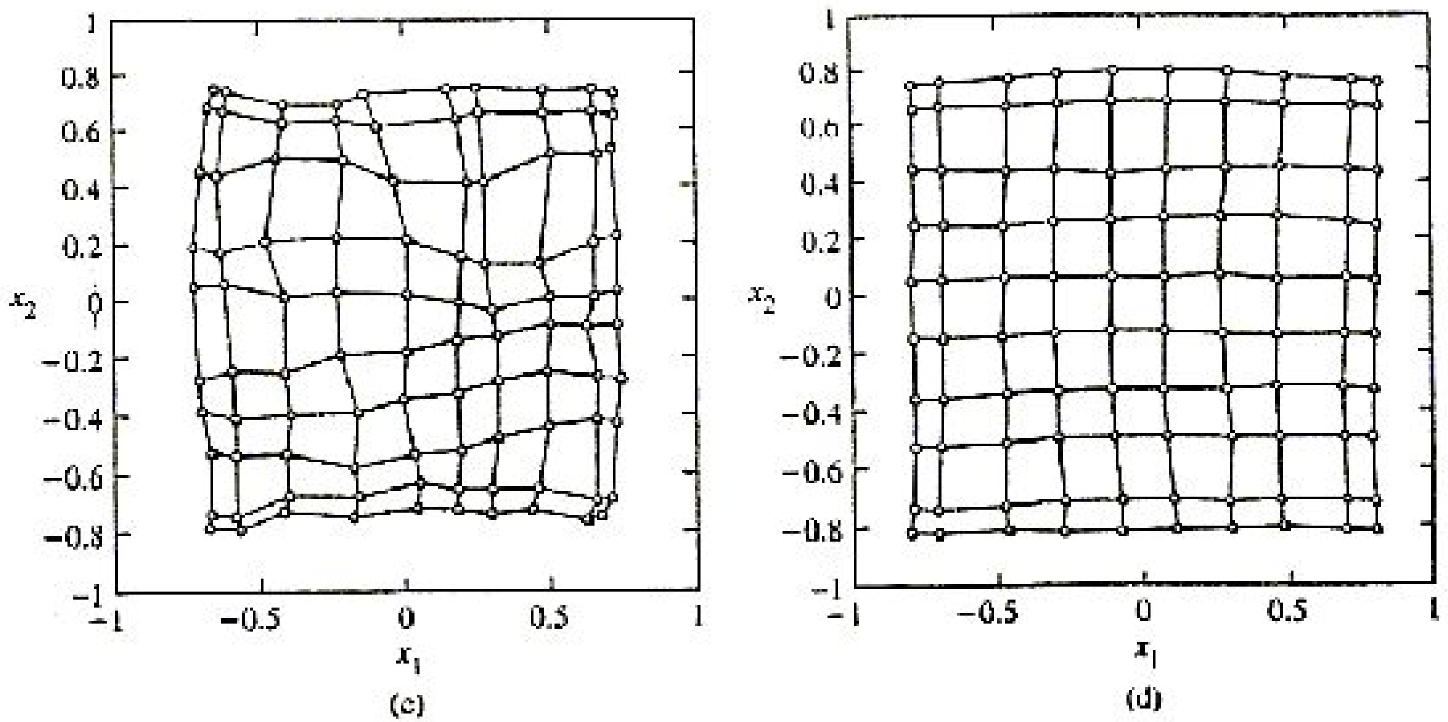
Two-dimensional input space, uniformly distributed

Output space is a two-dimensional lattice

a) **Initial random weights**

b) **After 50 iterations**

Ordering Process (II)



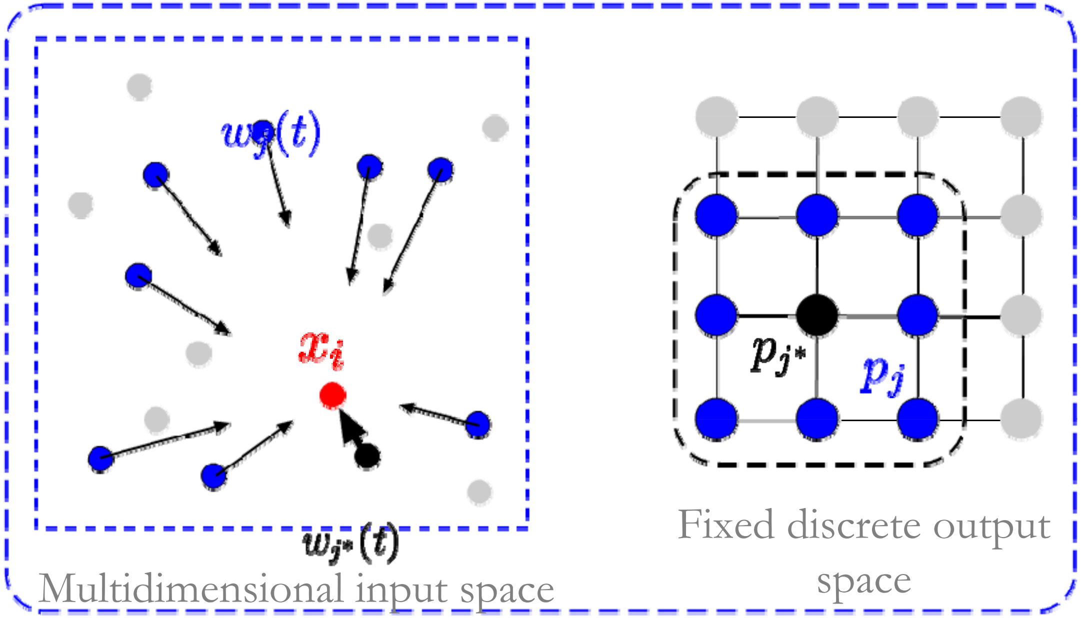
c) **Network after 1000 iterations**

d) **After 10,000 iterations**

SOM Algorithm

- Initialize the prototypes (codebook vectors)
- Present the samples
  - Search for the nearest prototype (most similar)
  - Update the nearest prototypes and its neighbors so that they become more similar to the sample
  - Decrease the range of neighborhood and the value of the learning rate a little
- Repeat until a stop criterion is satisfied

SOM Algorithm (II)



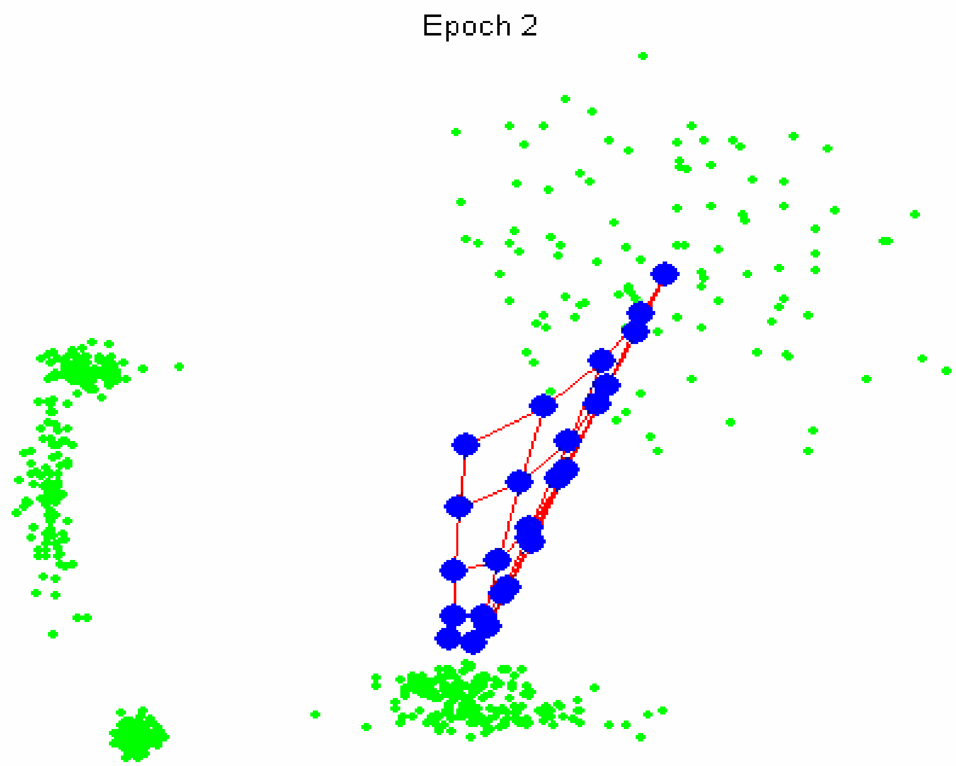
The winner and its neighbors in Nc are updated

$$j^* = \text{argmin}_{j=1\ldots N} ||x_i - w_j||$$

$$\Delta w_j(t) = w_j(t+1) - w_j(t) = \alpha(t) h_{jj^*}(t)(x_i - w_j(t))$$

$$h_{jj^*}(t) = e^{-\frac{\| p_j - p_j^* \|^2}{L(t)^2}} \qquad \alpha(t) = \alpha_0 \left(\frac{\alpha_f}{\alpha_0}\right)^{\left(\frac{t}{t_{max}}\right)} \qquad L(t) = L_0 \left(\frac{L_f}{L_0}\right)^{\left(\frac{t}{t_{max}}\right)}$$

Example: Kohonen’s Map in 2D



- It uses a 2D output grid for visualization of high-dimensional data

U-Matrix and Kohonen Maps

- **Although SOM has a topology preserving output space, this space is fixed and does not preserve distances**
- **Distance between codebook vectors is not represented in the maps**
- **Usually post-processing techniques such as U-matrix are required to visualize cluster boundaries in SOM**
- **U-Matrix: The sum of distances to its neighbors is displayed as elevation at the position of each neuron (3D) or using gray scale or colors (dark shade represent large distances)**

Example: Poverty Map

- Statistics published by World Bank, with 39 indicator values describing the poverty of the countries
- Only 78 out of 126 countries had 12 or more features out of the 39 possible (missing data)
- The data of these 78 countries was used to train a SOM (capital letters)
- The remaining 48 countries were labeled after training (lowercase letters)

World Poverty Map

Table 3.5. Legend of symbols used in Figs. 3.23 and 3.24:

AFG	Afghanistan	GRC	Greece	NOR	Norway
AGO	Angola	GTM	Guatemala	NPL	Nepal
ALB	Albania	HKG	Hong Kong	NZL	New Zealand
ARE	United Arab Emirates	HND	Honduras	OAN	Taiwan, China
ARG	Argentina	HTI	Haiti	OMN	Oman
AUS	Australia	HUN	Hungary	PAK	Pakistan
AUT	Austria	HVO	Burkina Faso	PAN	Panama
BDI	Burundi	IDN	Indonesia	PER	Peru
BEL	Belgium	IND	India	PHL	Philippines
BEN	Benin	IRL	Ireland	PNG	Papua New Guinea
BGD	Bangladesh	IRN	Iran, Islamic Rep.	POL	Poland
BGR	Bulgaria	IRQ	Iraq	PRT	Portugal
BOL	Bolivia	ISR	Israel	PRY	Paraguay
BRA	Brazil	ITA	Italy	ROM	Romania
BTN	Bhutan	JAM	Jamaica	RWA	Rwanda
BUR	Myanmar	JOR	Jordan	SAU	Saudi Arabia
BWA	Botswana	JPN	Japan	SDN	Sudan
CAF	Central African Rep.	KEN	Kenya	SEN	Senegal
CAN	Canada	KHM	Cambodia	SGP	Singapore
CHE	Switzerland	KOR	Korea, Rep.	SLE	Sierra Leone
CHL	Chile	KWT	Kuwait	SLV	El Salvador
CHN	China	LAO	Lao PDR	SOM	Somalia
CIV	Cote d'Ivoire	LBN	Lebanon	SWE	Sweden
CMR	Cameroon	LBR	Liberia	SYR	Syrian Arab Rep.
COG	Congo	LBY	Libya	TCD	Chad
COL	Colombia	LKA	Sri Lanka	TGO	Togo
CRI	Costa Rica	LSO	Lesotho	THA	Thailand
CSK	Czechoslovakia	MAR	Morocco	TTO	Trinidad and Tobago
DEU	Germany	MDG	Madagascar	TUN	Tunisia
DNK	Denmark	MEX	Mexico	TUR	Turkey
DOM	Dominican Rep.	MLI	Mali	TZA	Tanzania
DZA	Algeria	MNG	Mongolia	UGA	Uganda
ECU	Ecuador	MOZ	Mozambique	URY	Uruguay
EGY	Egypt, Arab Rep.	MRT	Mauritania	USA	United States
ESP	Spain	MUS	Mauritius	VEN	Venezuela
ETH	Ethiopia	MWI	Malawi	VNM	Viet Nam
FIN	Finland	MYS	Malaysia	YEM	Yemen, Rep.
FRA	France	NAM	Namibia	YUG	Yugoslavia
GAB	Gabon	NER	Niger	ZAF	South Africa
GBR	United Kingdom	NGA	Nigeria	ZAR	Zaire
GHA	Ghana	NIC	Nigaragua	ZMB	Zambia
GIN	Guinea	NLD	Netherlands	ZWE	Zimbabwe

World Poverty Map

BEL	SWE	ITA	YUG	rom	-	CHN TUR	bur IDN	MDG	-	BGD NPL	btn	afg gin MLI ner SLE
AUT che DEU FRA	NLD	JPN	-	bgr csk	HUN POL PRT	-	-	-	gab lbr	khm	PAK	moz mrt sdn yem
-	-	ESP	GRC	-	-	THA	MAR	-	IND	caf	SEN	MWI TZA uga
DNK GBR NOR	FIN	IRL	-	URY	ARG	ECU mex	-	EGY	hti	lao png ZAR	-	tcd
-	-	-	KOR	-	zaf	-	TUN	dza irq	GHA	NGA	-	ETH
CAN USA	-	ISR	-	-	COL PER	lbn	lby	ZWE	omn	-	ago	hvo
-	AUS	-	MUS tto	-	-	IRN PRY syr	hnd	BWA	KEN	BEN CIV	cog som	bdi RWA
NZL	-	-	CHL	PAN	alb	mng sau	-	vnm	jor nic	-	-	tgo
-	HKG SGP	are	CRI VEN	kwt	JAM MYS	-	DOM LKA PHL	-	BOL BRA SLV	-	GTM	CMR lso nam ZMB

Fig. 3.23. "Poverty map" of 126 countries of the world. The symbols written in capital letters correspond to countries used in the formation of the map; the rest of the symbols (in lower case) signify countries for which more than 11 attribute values were missing, and which were mapped to this SOM after the learning phase.

U-Matrix

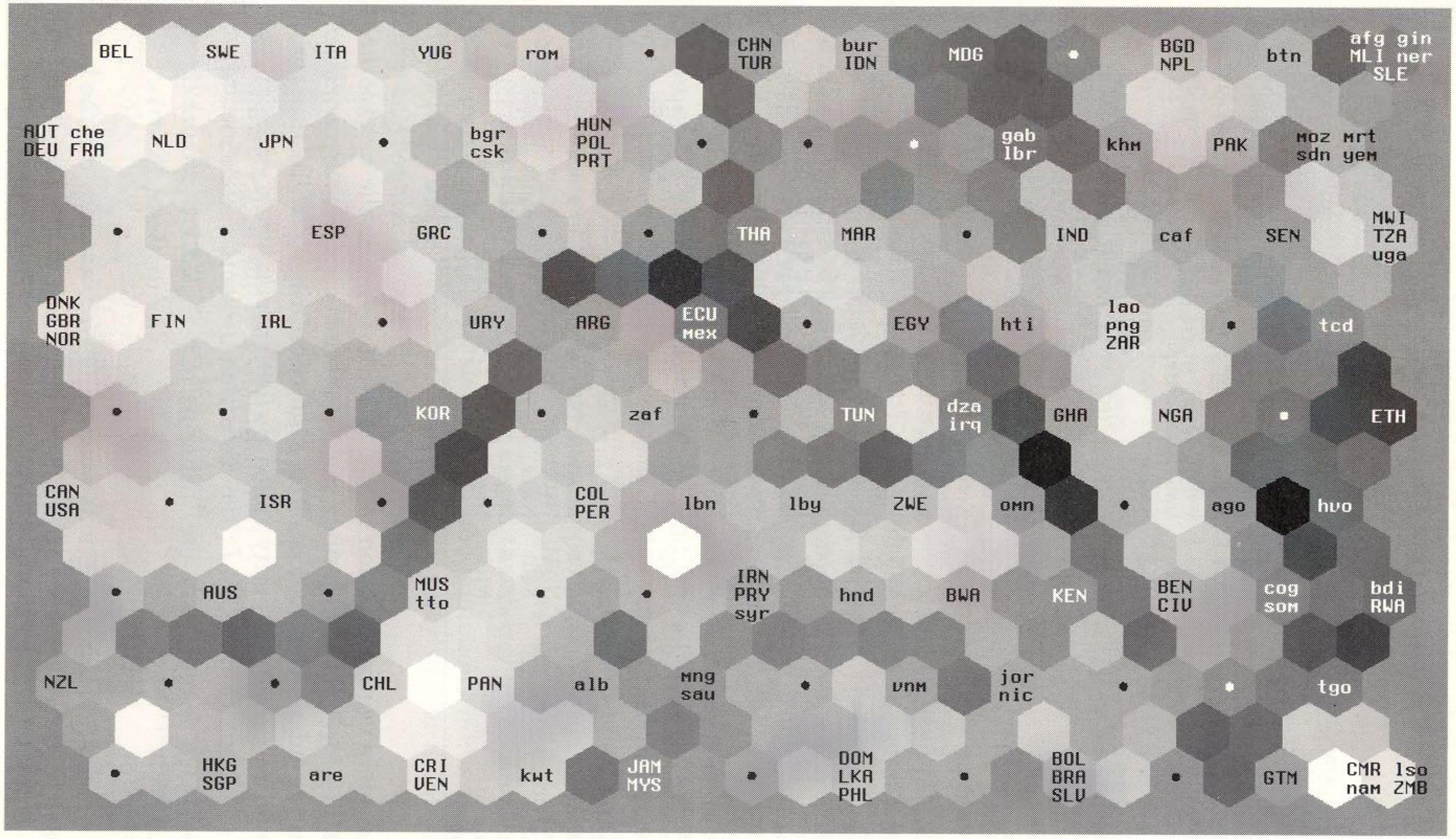
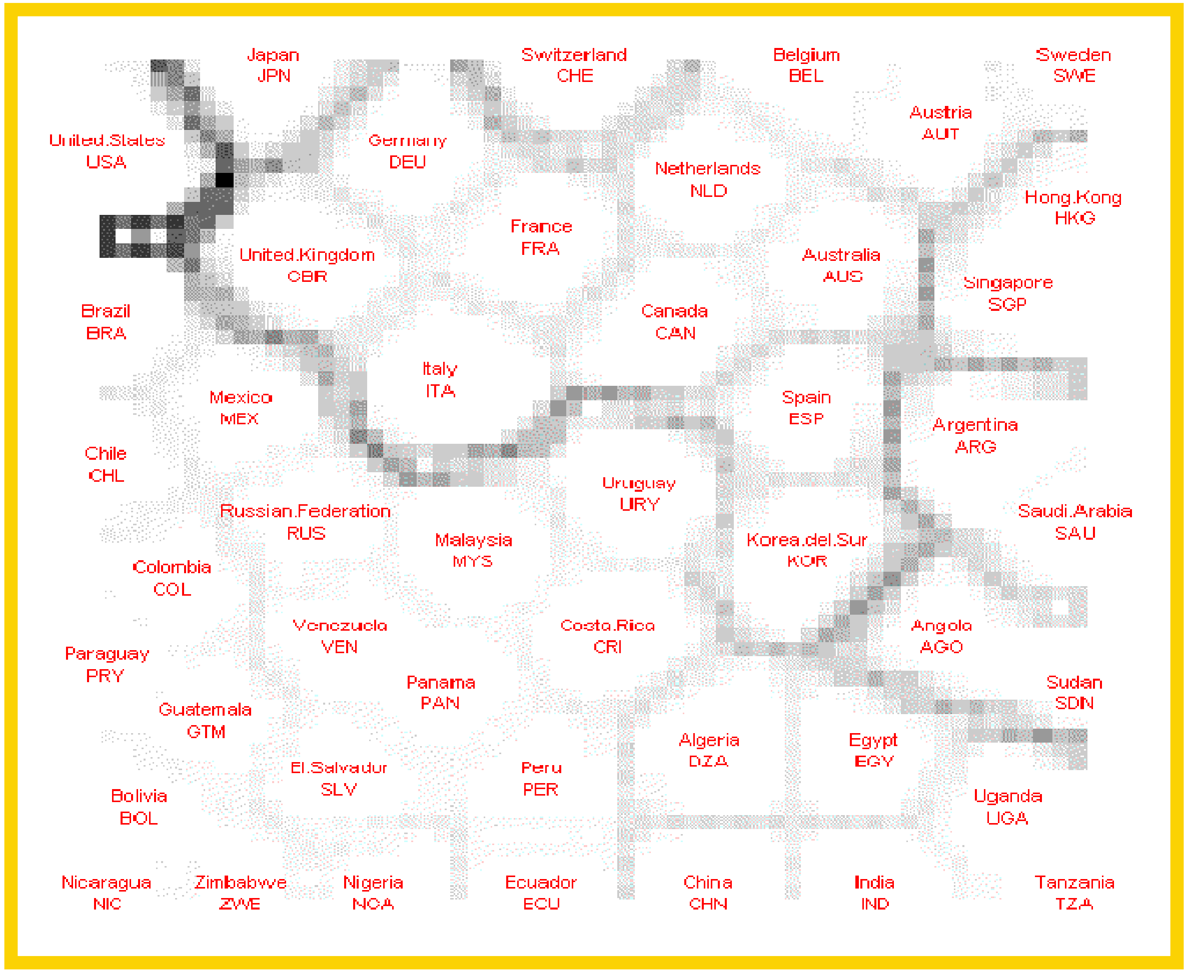


Fig. 3.24. "Poverty map," with the clustering shown by shades of gray

World Map using SOM and U-Matrix

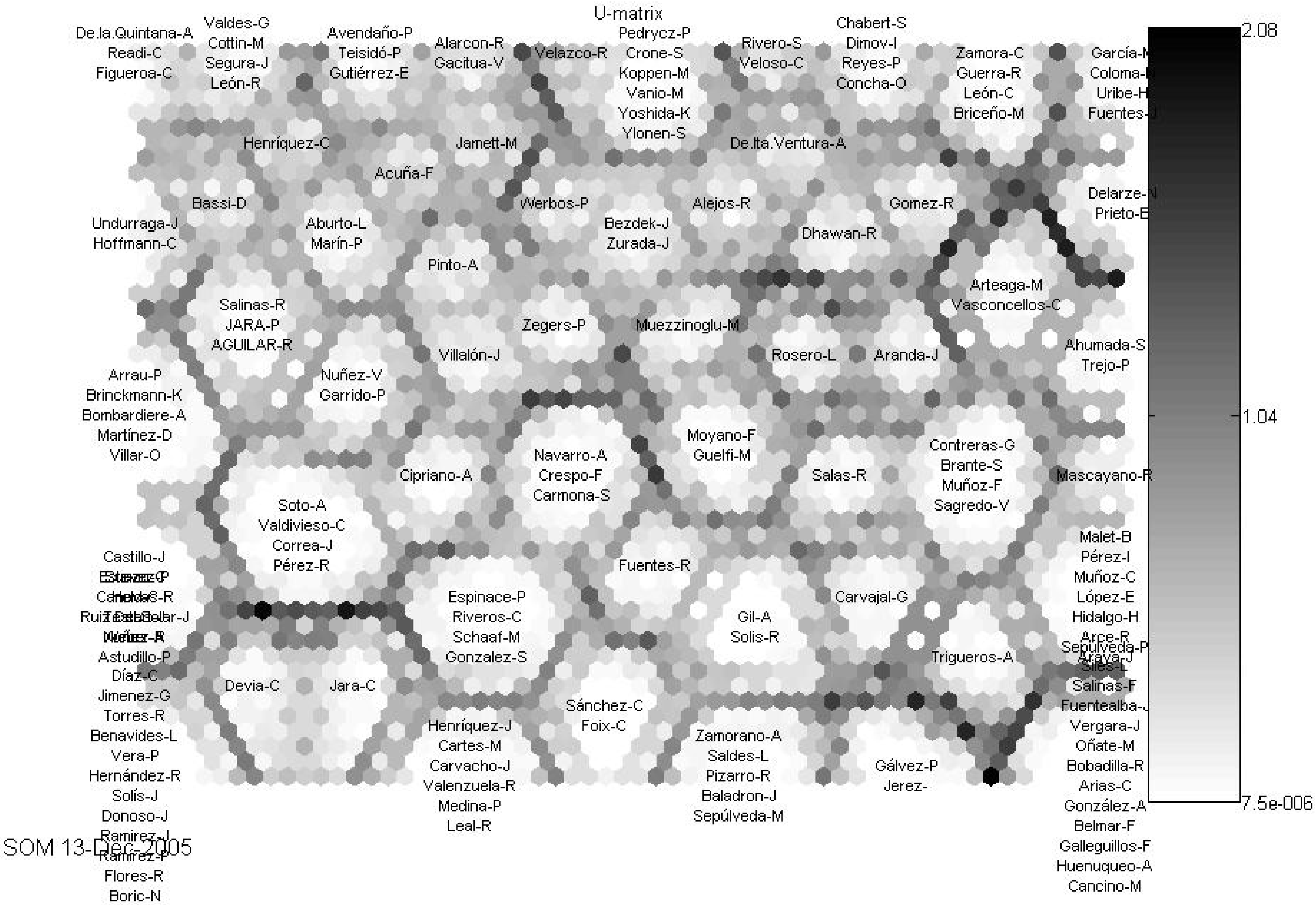


- Kohonen’s map using 4 economic variables (1999)

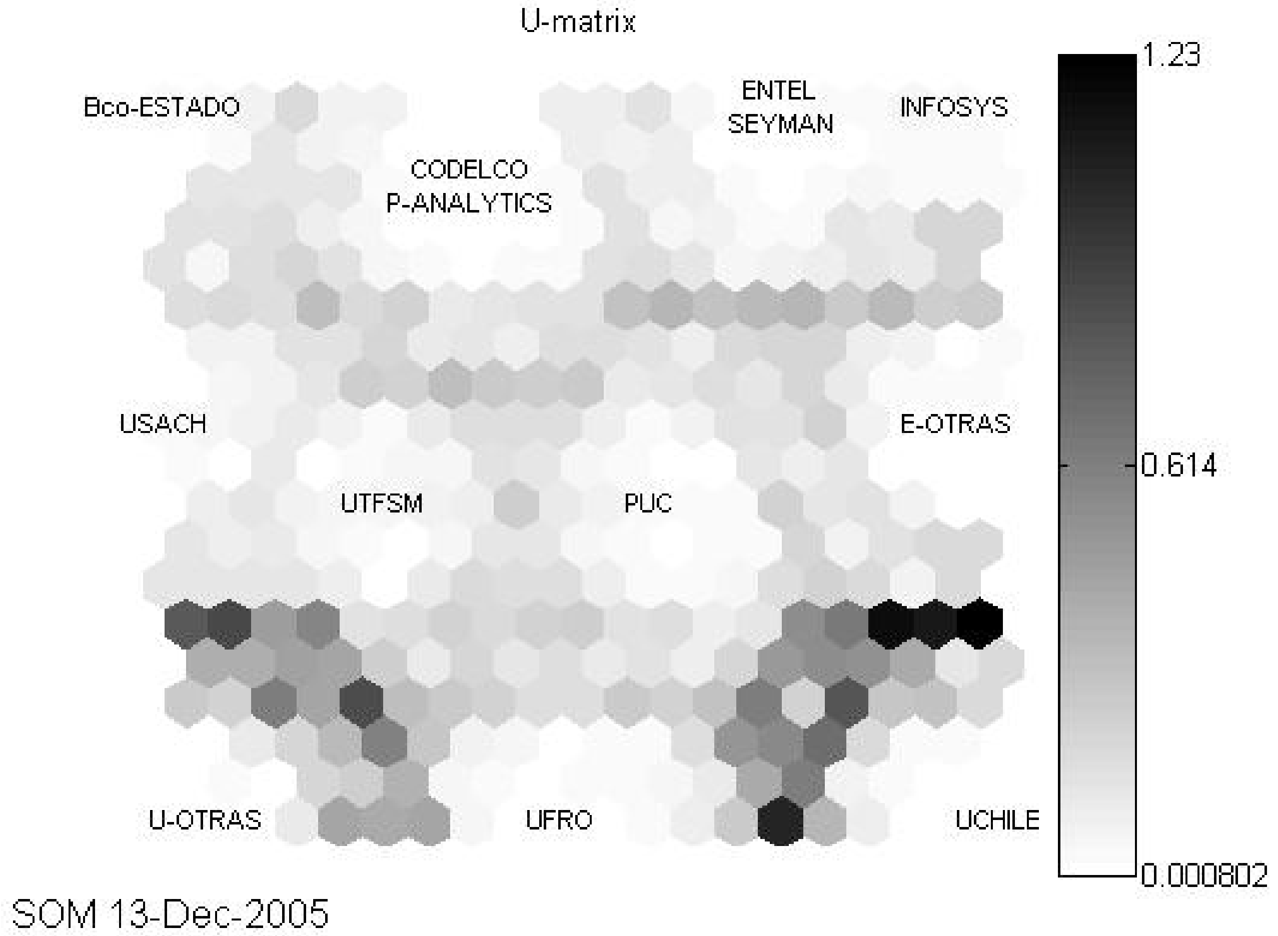
Example: EVIC 2005 Audience

- The following binary features were defined:
  - "University“: UCHILE UTFSM UFRO PUC USACH OTHERS
  - "Industry“: ENTEL B_ESTADO CODELCO P_ANALYTICS INFOSYSSEYMAN OTHERS
  - "Sex“: FEMALE MALE
  - “Student" : YES NO
  - "IEEE Membership“ YES NO
  - “Country“: AU CL CO MX UY PE USA OTHER
  - "Region“: RM V VII VIII IX

U-Matrix


U-matrix
De.la.Quintana-A
Readi-C
Figueroa-C
Valdes-G
Cottin-M
Segura-J
León-R
Avendaño-P
Teisidó-P
Gutiérrez-E
Alarcon-R
Gacitua-V
Velazco-R
Pedrycz-P
Crone-S
Koppen-M
Vanio-M
Yoshida-K
Ylonen-S
Rivero-S
Veloso-C
Chabert-S
Dimov-I
Reyes-P
Concha-O
Zamora-C
Guerra-R
León-C
Briceño-M
Henríquez-C
Jamett-M
De.Ita.Ventura-A
Acuña-F
Bassi-D
Werbos-P
Alejos-R
Gomez-R
Undurraga-J
Hoffmann-C
Aburto-L
Marín-P
Bezdek-J
Zurada-J
Dhawan-R
Pinto-A
Arteaga-M
Vasconcellos-C
Salinas-R
JARA-P
AGUILAR-R
Zegers-P
Muezzinoglu-M
Ahumada-S
Trejo-P
Villalón-J
Rosero-L
Aranda-J
Arrau-P
Brinckmann-K
Bombardiere-A
Martínez-D
Villar-O
Nuñez-V
Garrido-P
Moyano-F
Guelfi-M
Contreras-G
Brante-S
Muñoz-F
Sagredo-V
Navarro-A
Crespo-F
Carmona-S
Cipriano-A
Salas-R
Soto-A
Valdivieso-C
Correa-J
Pérez-R
Malet-B
Pérez-I
Muñoz-C
López-E
Hidalgo-H
Castillo-J
Fuentes-R
Espinace-P
Riveros-C
Schaaf-M
Gonzalez-S
Gil-A
Solis-R
Carvajal-G
Trigueros-A
Astudillo-P
Díaz-C
Jimenez-G
Torres-R
Benavides-L
Vera-P
Hernández-R
Solís-J
Donoso-J
Flores-R
Boric-N
Devia-C
Jara-C
Sánchez-C
Foix-C
Henríquez-J
Cartes-M
Carvacho-J
Valenzuela-R
Medina-P
Leal-R
Zamorano-A
Saldes-L
Pizarro-R
Baladron-J
Sepúlveda-M
Gálvez-P
Jerez-
Salinas-F
Fuentealba-J
Vergara-J
Oñate-M
Bobadilla-R
Arias-C
González-A
Belmar-F
Galleguillos-F
Huenuqueo-A
Cancino-M
2.08
1.04
7.5e-006
SOM 13-Dec-2005

Kohonen´s Map of EVIC Audience



Neural Gas (NG)

- NG does not define an output space, in contrast to SOM
- NG updates the codebook vectors using a neighborhood ranking function in the input space, instead of a neighborhood in output space
- NG minimizes a global cost function, while no cost function exists for the SOM adaptation rule
- NG can outperform SOM when quantizing topologically arbitrary manifolds (since it is not constrained by an output grid)
- Only a few visualization schemes have been developed for NG

Neural Gas (NG) Algorithm

- Initialize codebook vectors w_j randomly
- Present input vector $x_i(t)$ at iteration t
- Find the nearest codebook vector

  and generate a ranking $r(x_i,w_j)$ for each w_j with respect to the input vector x_i $j^* = \mathrm{argmin}_{j=1 \ldots N} \|x_i - w_j\|$
- Update the codebook vectors

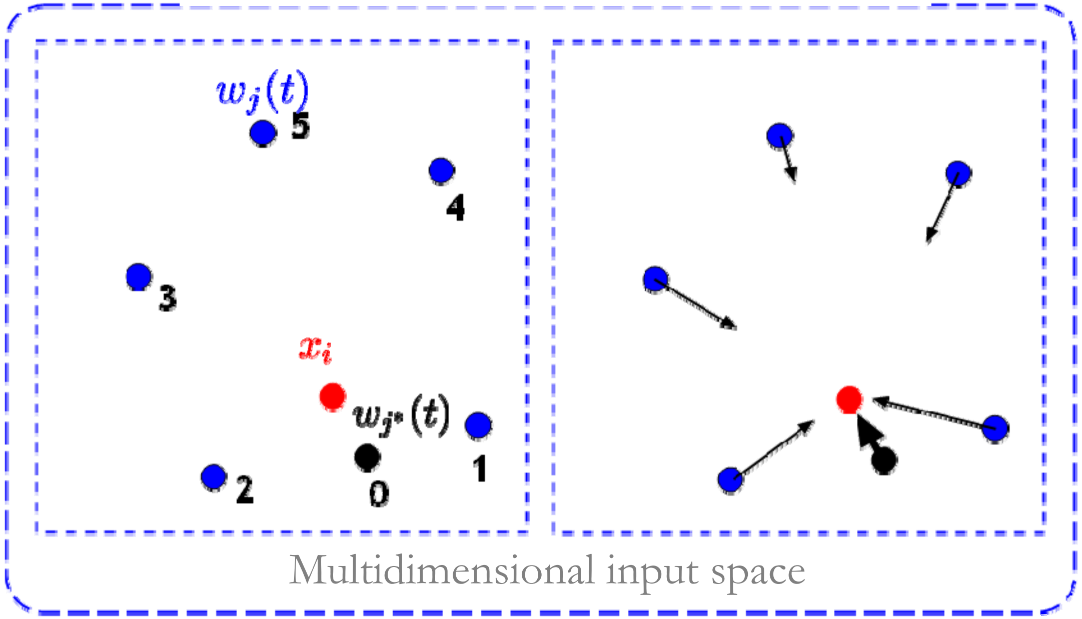
  where is the neighborhood function

$$\Delta w_j(t) = w_j(t+1) - w_j(t) = \epsilon(t) h_\lambda(t)(x_i - w_j(t))$$

$$h_\lambda(x_i, w_j) = e^{-\frac{r(x_i, w_j)}{\lambda(t)}}$$

Neural Gas (NG)

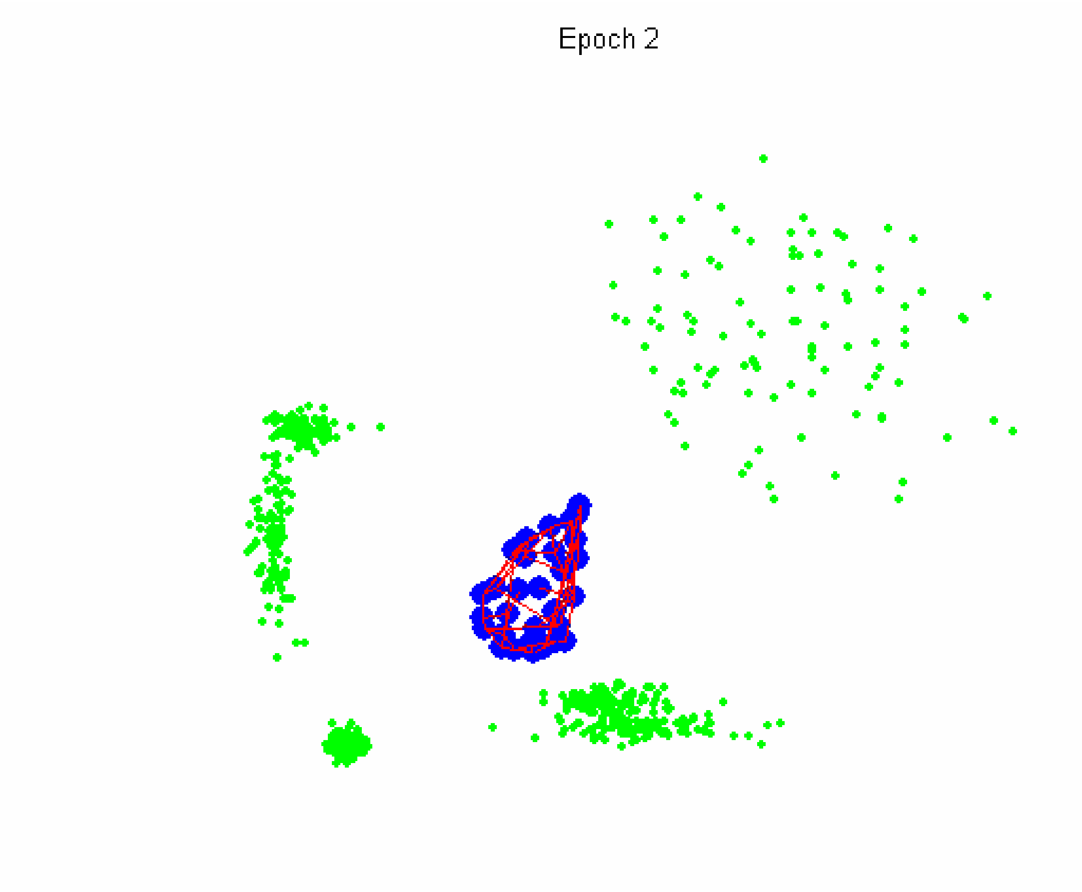
$$j^* = \mathrm{argmin}_{j=1\ldots N} \|x_i - w_j\|$$



$$\Delta w_j(t) = w_j(t+1) - w_j(t) = \epsilon(t) h_\lambda(t)(x_i - w_j(t))$$

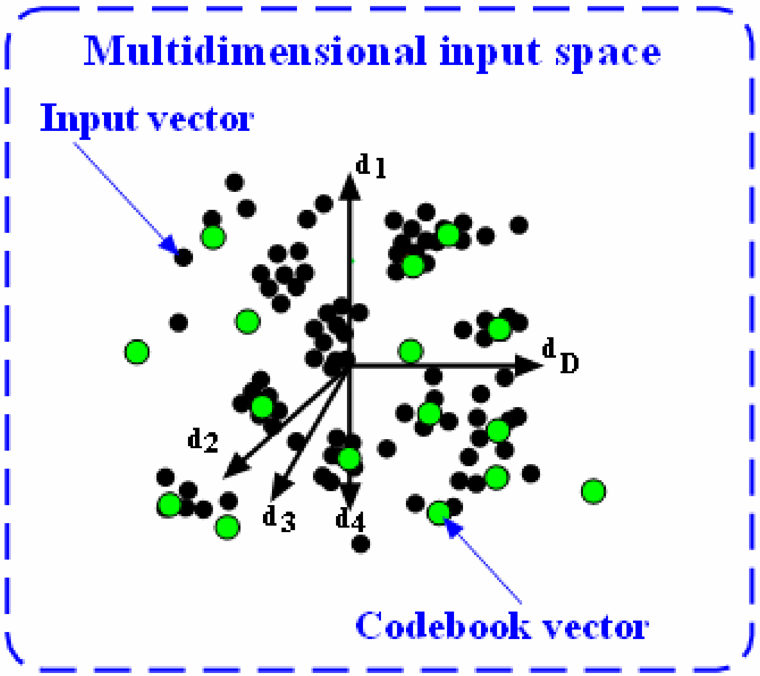
$$h_\lambda(x_i, w_j) = e^{-\frac{R_{ij}}{\lambda(t)}} \qquad \epsilon(t) = \epsilon_0 \left(\frac{\epsilon_f}{\epsilon_0}\right)^{\left(\frac{t}{t_{max}}\right)} \qquad \lambda(t) = \lambda_0 \left(\frac{\lambda_f}{\lambda_0}\right)^{\left(\frac{t}{t_{max}}\right)}$$

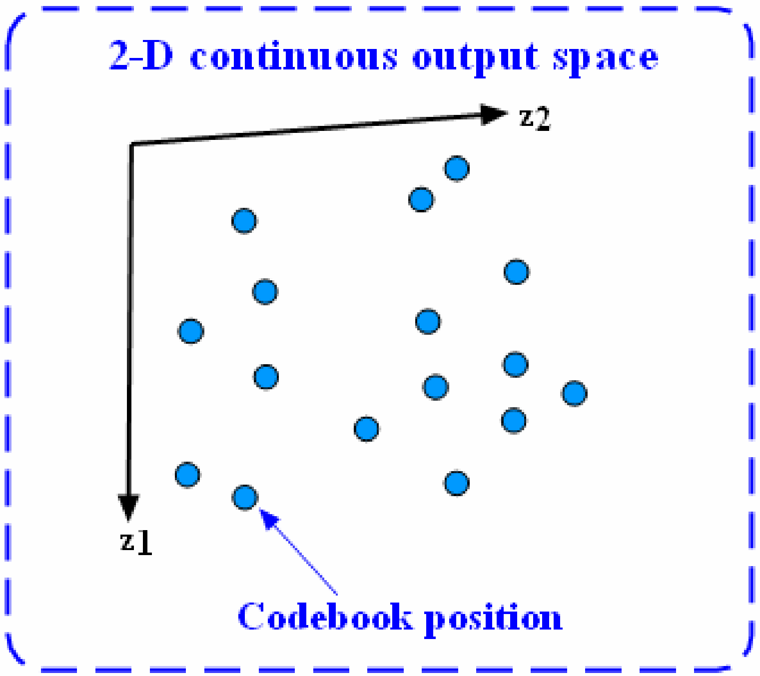
Example of Neural Gas



- Connections are made between the best matching unit and the second closest
- Connections are allowed to age and are removed eventually if not refreshed

Notation definition





$$d_{j,k} = \|\mathbf{w}_j - \mathbf{w}_k\| = \sqrt{\sum_{d=1}^{D}(w_{j,d} - w_{k,d})^2}$$

Interpoint distance in input space

$$D_{j,k} = \|\mathbf{z}_j - \mathbf{z}_k\| = \sqrt{\sum_{a=1}^{A}(z_{j,a} - z_{k,a})^2}$$

Interpoint distance in output space

Multi-Dimensional Scaling (MDS)

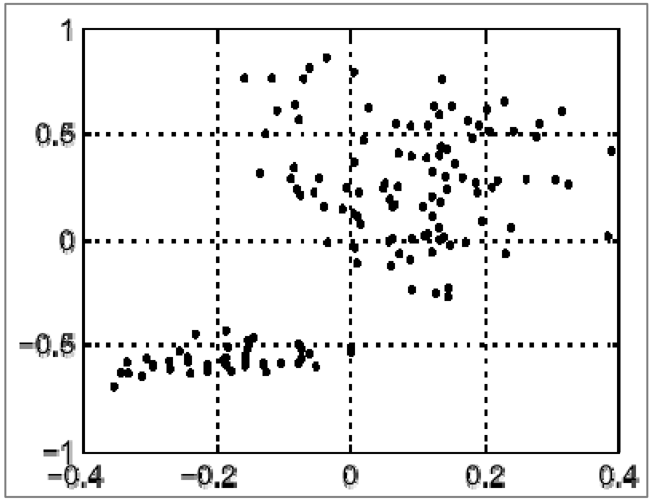
- The aim of MDS is to provide a visual representation of the pattern of dissimilarities among a set of objects, such that the distances between points in the map match the original dissimilarities between pairs of objects.
  - Togerson's Classical Scaling
    - Dissimilarities are Euclidean distances
    - Spectral matrix decomposition
  - Sammon's Non-linear Mapping (NLM)
    - It allows a monotonic transformation of dissimilarities
    - It minimizes a loss function by steepest descent
- Computational load $O(N^2)$

Sammon stress (error)

$$E = \frac{1}{\sum_{j=1}^{N}\sum_{i=1}^{j} d_{ij}} \sum_{j=1}^{N}\sum_{i=1}^{j} \frac{[d_{ij} - D_{ij}]^2}{d_{ij}}$$

dij is the distance between the i-th and the j-th points in input space,
Dij is the distance between the i-th and the j-th points in output space
N number of points to be projected

- Quadratic optimization strategy
- Prone to get stuck at local minima



Iris 2-D projection using NLM

NLM (Sammon) Algorithm

1. Compute the matrix of pair-wise distances in input space

$$d_{ij} = \left[\sum_{\ell=1}^{D} \left(x_{i\ell} - x_{j\ell} \right)^2 \right]^{\frac{1}{2}}$$

   where x_i and x_j a … ectors

2. Initialize positior … ce

$$y_{i\ell}(0)$$

3. Compute the matrix of pair-wise distances in output space

   where y_i and y_j are A-dimensional output vectors

4. Compute Sammon error E

$$D_{ij} = \left[\sum_{\ell=1}^{A} \left(y_{i\ell} - y_{j\ell} \right)^2 \right]^{\frac{1}{2}}$$

NLM Algorithm(II)

5. Update vector positions in output space

$$y_{i\ell}(t+1) = y_{i\ell}(t) - \xi \frac{\partial E(t)/\partial y_{i\ell}(t)}{\left|\partial^2 E(t)/\partial y_{i\ell}(t)^2\right|} \qquad 0 < \xi \le 1$$

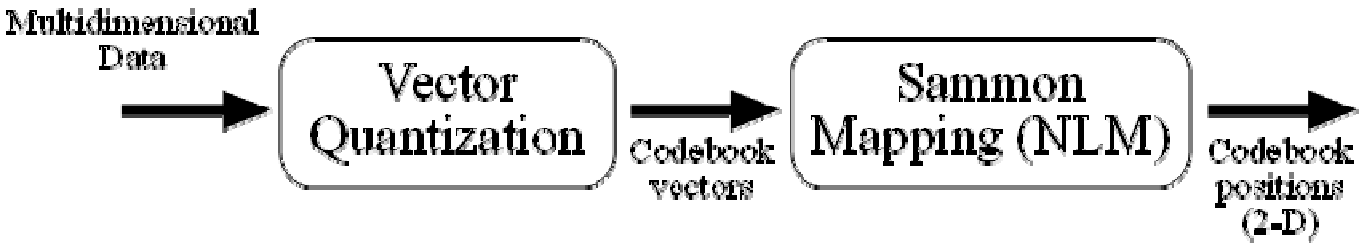
where

$$\partial E(t)/\partial y_{i\ell}(t) = -\frac{2}{c}\sum_{k \ne i}^{N}\left(y_{i\ell}(t) - y_{k\ell}(t)\right) \times \frac{d_{ik} - D_{ik}(t)}{d_{ik} D_{ik}(t)}$$

6. $$\partial^2 E(t)/\partial y_{i\ell}(t)^2 = -\frac{2}{c}\sum_{k \ne i}^{N}\frac{1}{d_{ik} D_{ik}(t)}\left(\left(d_{ik} - D_{ik}(t)\right) - \frac{\left(y_{i\ell}(t) - y_{k\ell}(t)\right)^2}{D_{ik}(t)}\left(1 + \frac{d_{ik} - D_{ik}(t)}{D_{ik}(t)}\right)\right)$$

$$c = \sum_{j=1}^{N}\sum_{i=1}^{j} d_{ij}$$

Vector Quantization + NLM



- Off-line visualization scheme that combines a vector quantizer with NLM
  - Self-organizing map (SOM)
  - Neural Gas (NG)
- Each codebook vector in input space has associated a codebook position in output space

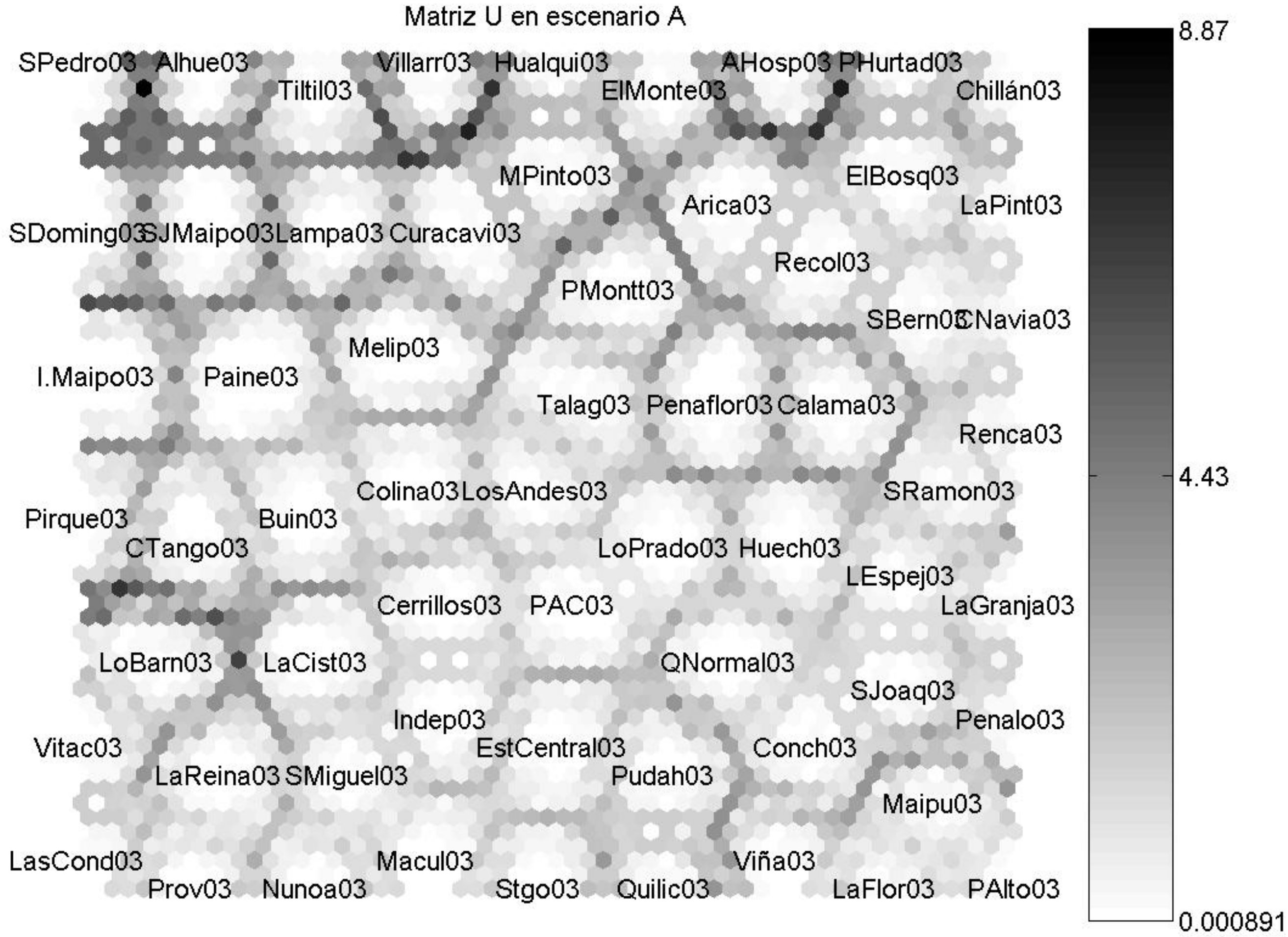
Example: CASEN survey

- The CASEN survey delivers several social indicators at the county, region and country levels in Chile.
- It aims at making a diagnosis and to evaluate the impact of the government social policy
- Data is publicly available at www.mideplan.cl/casen/
- Each county is described by 36 attributes, including income, social security, electricity and water supplies, etc.
- There are 52 counties in the metropolitan region of Santiago

List of Attributes

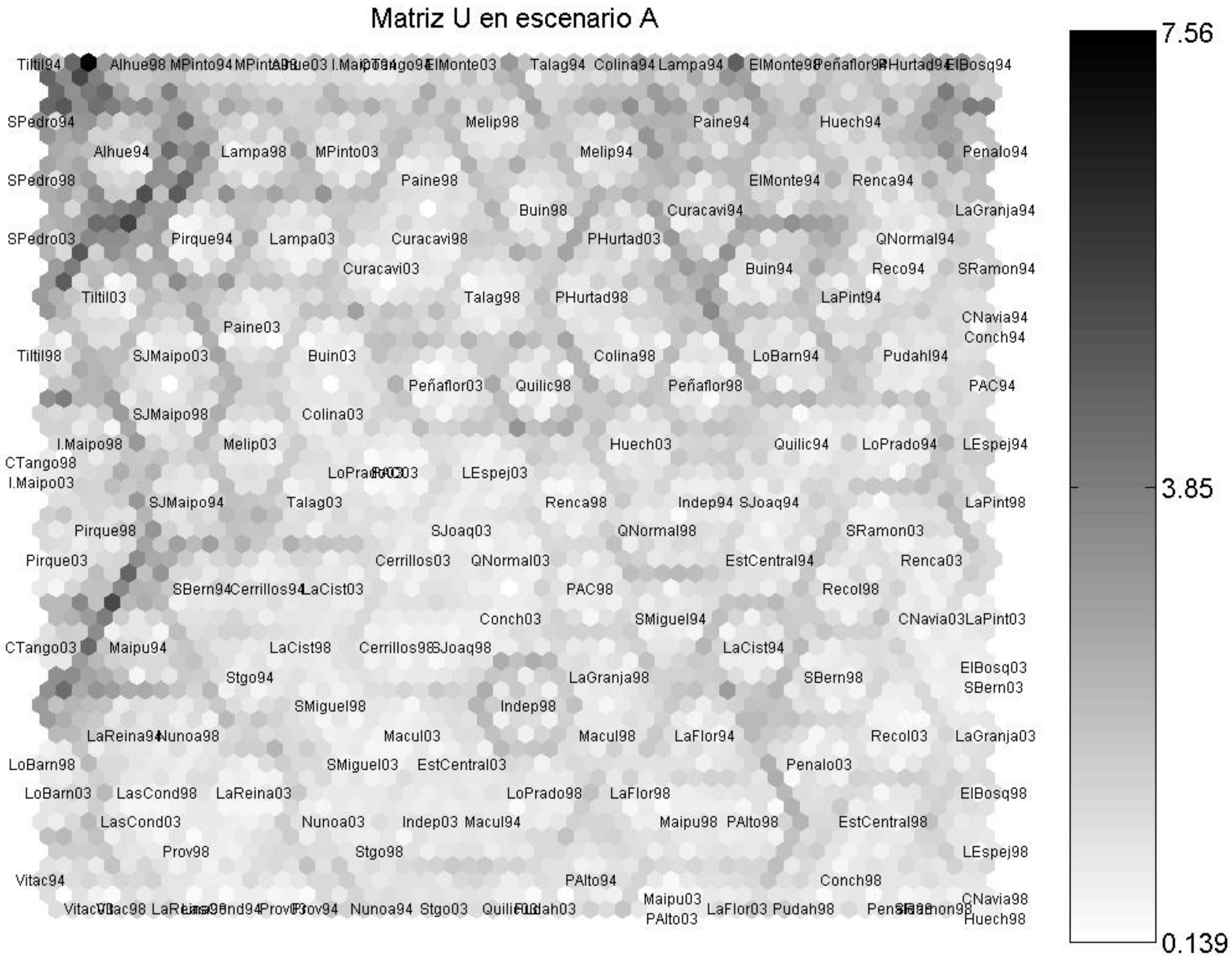
Índice	Descripción
1	Nombre de la comuna
2	Población
3	Agua Potable red pública
4	Agua Potable otro
5	Electricidad con red pública
6	Electricidad sin red pública
7	Electricidad no dispone
8	Educación promedio escolaridad
9	Educación analfabetismo
10	Excretas WC conectado alcantarillado
11	Excretas WC conectado a fosa séptica
12	Excretas otro
13	Excretas no dispone
14	Empleo participación
15	Empleo ocupados
16	Empleo desocupación
17	Hacinamiento sin
18	Hacinamiento con
19	Viviendas buenas
20	Viviendas aceptables
21	Viviendas recuperables
22	Viviendas deficitarias
23	Saneamiento bueno
24	Saneamiento aceptable
25	Saneamiento regular
26	Saneamiento menos que regular
27	Saneamiento deficitarias
28	Ingreso autónomo (en U.F.)
29	Ingreso subsidios monetarios (en U.F.)
30	Ingreso monetario (en U.F.)
31	Hogares indigente
32	Hogares total pobres
33	Individuos indigente
34	Individuos total pobres
35	Previsión FONASA
36	Previsión ISAPRE
37	Previsión particular otros

SOM+U-Matrix



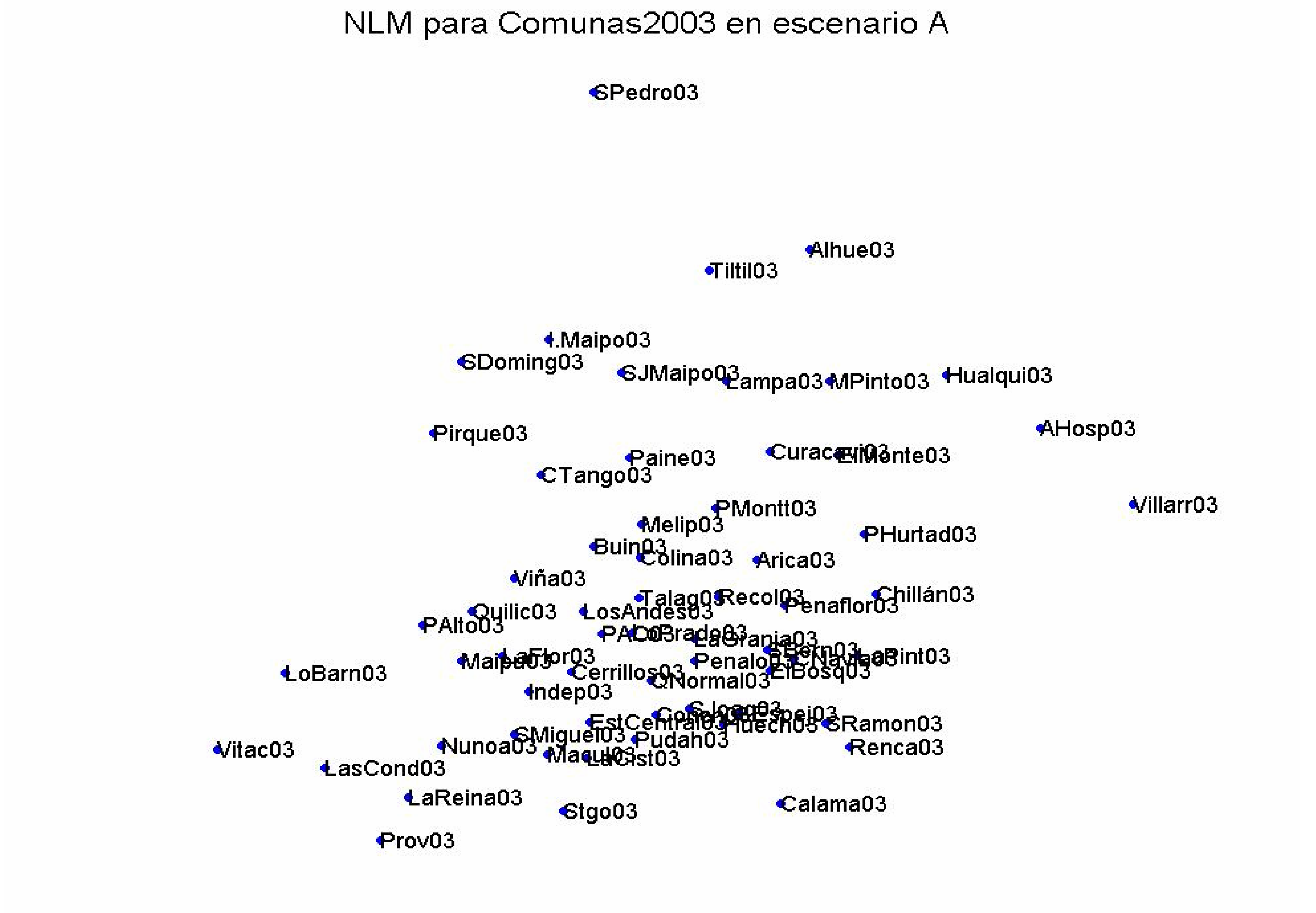
SOM para Comunas2003 usando grilla hexagonal de 30 x 30

SOM+U-Matrix

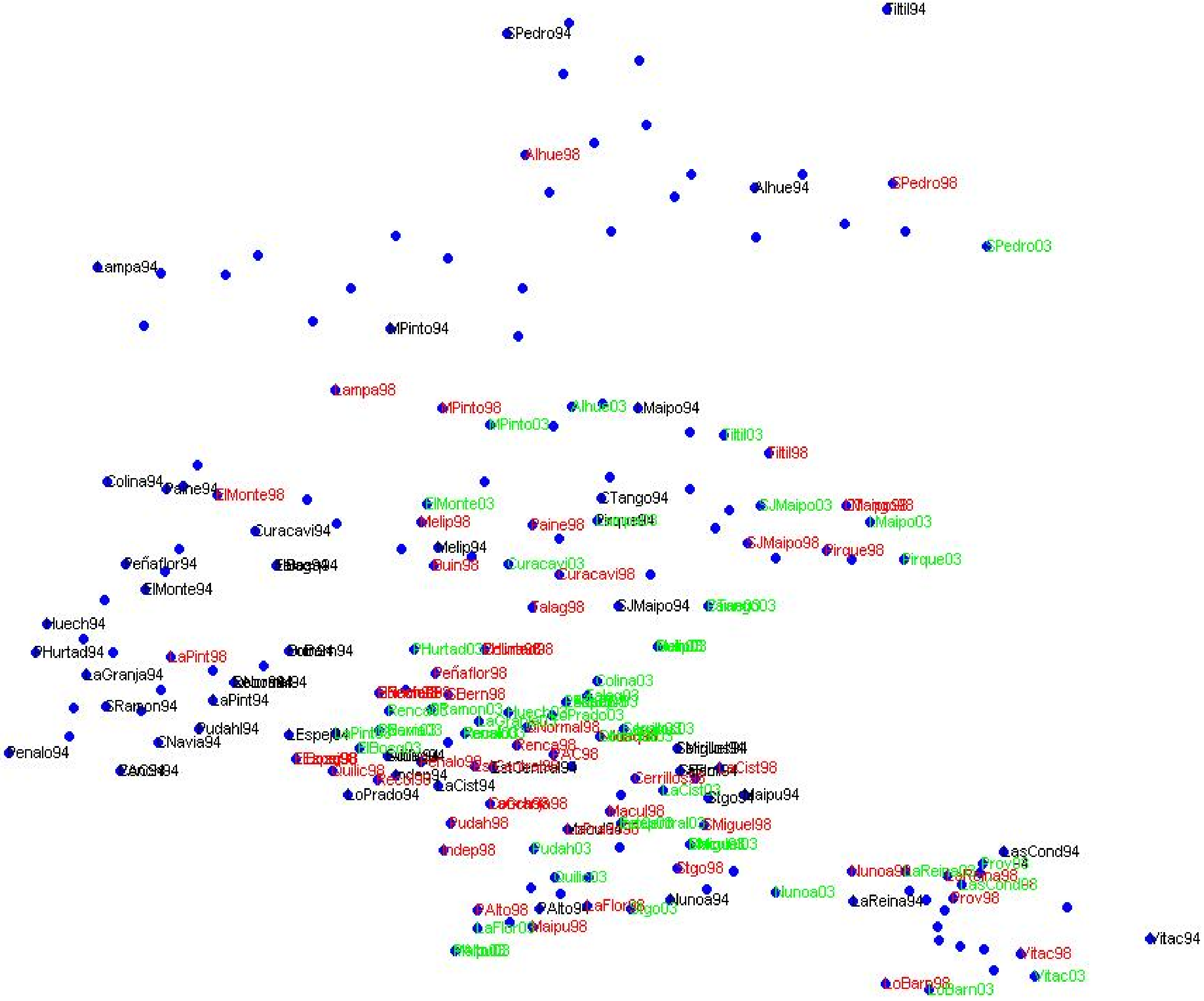


SOM para ComunasRMSecuencia usando grilla hexagonal de 30 x 30

Sammon Mapping (NLM)



NG/NLM para ComunasRMSecuencia en escenario A



SOM Toolbox Version 2.0

- Free software package for Matlab 5.0 (Compatible with Matlab 6.5 but not 7.0)
- http://www.cis.hut.fi/projects/somtoolbox
- It includes SOM, U-matrix, Neural Gas, Sammon´s Mapping, K-means, PCA, CCA, etc

Topology Preservation Measurement

- The topology preservation measurement q_m computes the rank order of the n nearest codebook vectors in input space and the n nearest codebook positions in output space
- $q_m=0$ indicates a poor neighborhood preservation, and $q_m=1$ indicates a perfect neighborhood preservation

$$q_m = \frac{1}{3n \times N} \sum_{j=1}^{N} \sum_{i=1}^{n} q_{m_{ji}}$$ where,

$$q_{m_{ji}} = \begin{cases} 3, & \text{if } NN_{ji^w} = NN_{ji^z} \\ 2, & \text{if } NN_{ji^w} = NN_{jl^z},\ l \in [1,n],\ i \neq l \\ 1, & \text{if } NN_{ji^w} = NN_{jt^z},\ t \in [n,k],\ n < k \\ 0, & \text{else.} \end{cases}$$

Trustworthiness and Continuity Measures

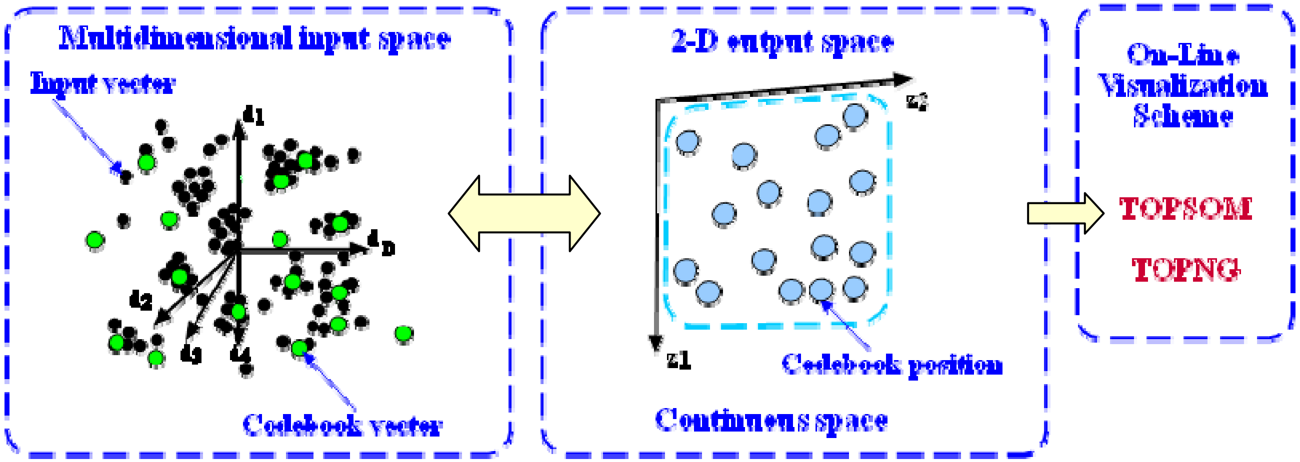
- A projection is said to be trustworthy if the set of k nearest neighbors of a point in the map are also close by in the original space
- A projection is said to be continuous if the set of k closest neighbors of a point in the original space are also close by in the output space
- These measures are mathematically defined

Visualization schemes based on NG

- Off-line
  - NG/NLM
  - NG-PSO (Particle Swarm Optimization) &
    Growing PSO
    (Figueroa, Estévez & Hernández, IJCNN 2005)
  - NG-CE (Cross Entropy)
    (Estévez, Figueroa & Saito, IJCNN 2005, Neural Networks Journal, 2005)
- On-line
  - TOPNG (Figueroa & Estévez, IJCNN 2004)
  - OVI-NG (Estévez & Figueroa, WSOM 2005; Neural Networks Special Issue on Self-Organizing Maps)

TOpological Projections (TOP)

Figueroa & Estévez, IJCNN 2004



- TOPSOM and TOPNG add a continuous output space
- Online visualization schemes: Codebook vectors and codebook positions are adjusted simultaneously
- Computational load is O(N)
- A heuristic rule is added for updating the codebook positions zi

$$z_j(t+1) = z_j(t) + \epsilon(t)(D_{j,j^*} - d_{j,j^*})(z_{j^*}(t) - z_j(t))$$

Online VIsualization Neural Gas

WSOM 2005, P. Estévez and C. Figueroa

Invited to Special Issue Neural Networks Journal

- A global cost function similar to the one used by CCA is introduced:

$$E = \frac{1}{2}\sum_{j=1}^{N}\sum_{k\neq j}(D_{j,k} - d_{j,k})^2 F(s_{j,k}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{k\neq j} E_{j,k}$$

- $F(s_{j,k})$ i vor local topolo
- A ranking of position vector in the output space wrt to the codebook position asociated to the winner, is introduced

$$s(y_i, z_j) \in \{0, 1, \ldots, N-1\}$$

OVI-NG (II)

- The weighting function F is an exponential function of the ranking of position vectors in the output space

$$F(f) = e^{-\frac{f}{\lambda_f}}$$

- Instead of using gradient des[illegible]ok position associated to the winner, $z_j^*(t)$, is fixed, and all the other positions are moved toward the winner’s position, disregarding any interaction among them
- Computational load is *O(NlogN)*

OVI-NG Learning Algorithm

1. Initialize codebook vectors, wj, and codebook positions, z_j, randomly. Set the connection set to the empty set.
2. Present an input vector, $x_i(t)$, to the network (i=1,…,M) at iteration t
3. Find the best matching unit (BMU):

   and generate $j^* = \mathrm{argmin}_{j=1\ldots N} ||x_i - w_j||$ ector $w_j(t)$ wrt the input vector $x_i(t)$.
4. Update the codebook vectors:

   where h_λ is the neighborhood ranking function.

$$\Delta w_j(t) = w_j(t+1) - w_j(t) = \epsilon(t) h_\lambda(t)(x_i - w_j(t))$$

OVI-NG Learning Algorithm (II)

5. If a connection between the BMU and the second closest unit does not exist already, create a connection between these units

where the index of $j^* - k^* : C_{in} = C_{in} \cup \{ w_j^*, w_k^* \}$

Set the edge of the connection j*-k* to zero (refresh the connection if the connections exist) $k^* = \arg\min_{i \in A \setminus j^*} \left\| x_i(t) - w_j(t) \right\|$

6. Increment the age of all edges emanating from w_j*:

$$age_{(j^*,k^*)} = 0$$

Remove those connections with an age exceeding the lifetime T(t).

$$age(j^*, \ell) = age(j^*, \ell) + 1$$

OVI-NG Learning Algorithm (III)

7. Generate the ranking $s(z_j^*(t), z_j(t))$ for each codebook position $z_j(t)$ wrt to the codebook position associated to the winner unit $z_j^*(t)$, with j* different than j.
8. Update the codebook positions:

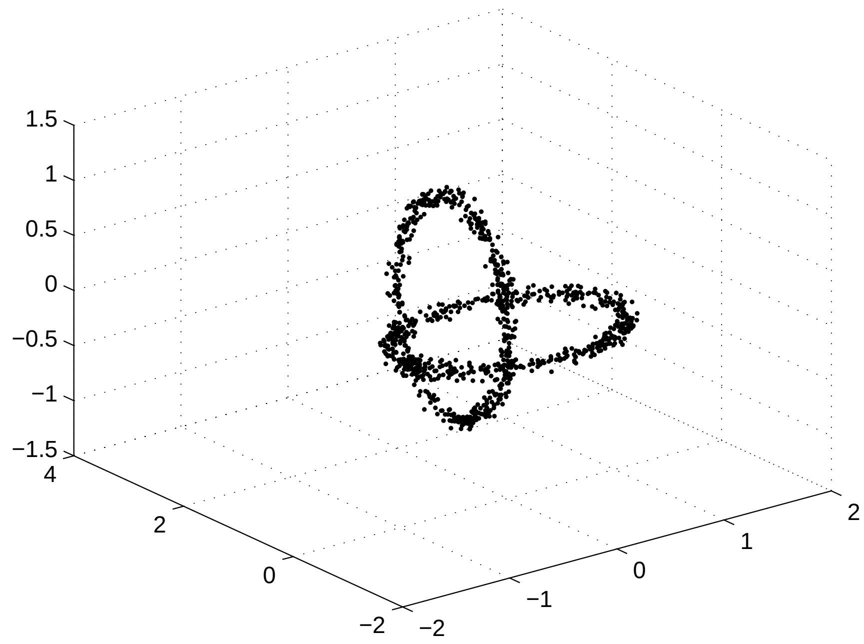
$$z_j(t+1) = z_j(t) + \alpha(t) F(s(z_j^*(t), z_j(t)) \frac{(D_{j,j^*} - d_{j,j^*})}{D_{j,j^*}} (z_{j^*}(t) - z_j(t))$$

9. If t<tmax go back to step 2)

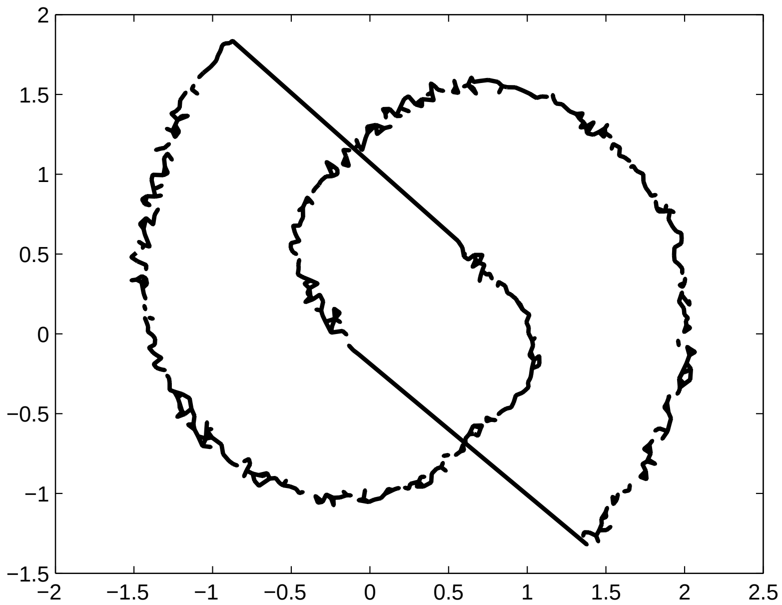
$$F(f) = e^{-\frac{f}{\lambda_f}}$$

Chainlink Data Set

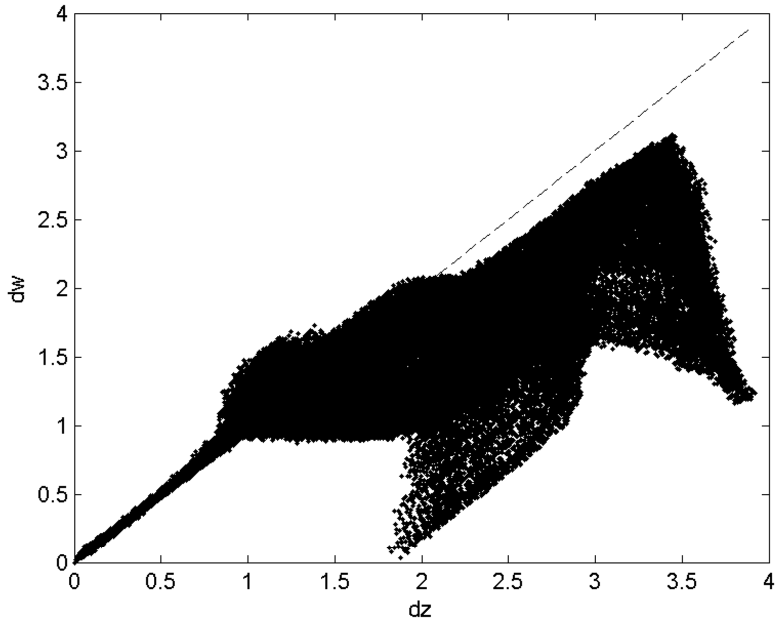
– 3 tore-shaped clusters of 500 points each, which are intertwined like the links of a chain



Chainlink Results

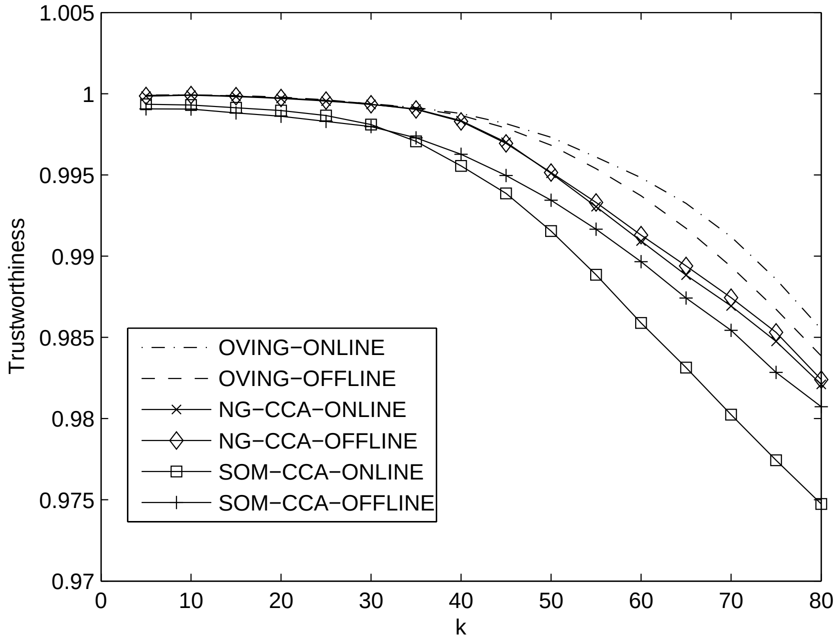


Quantized data projection
with CHR connections

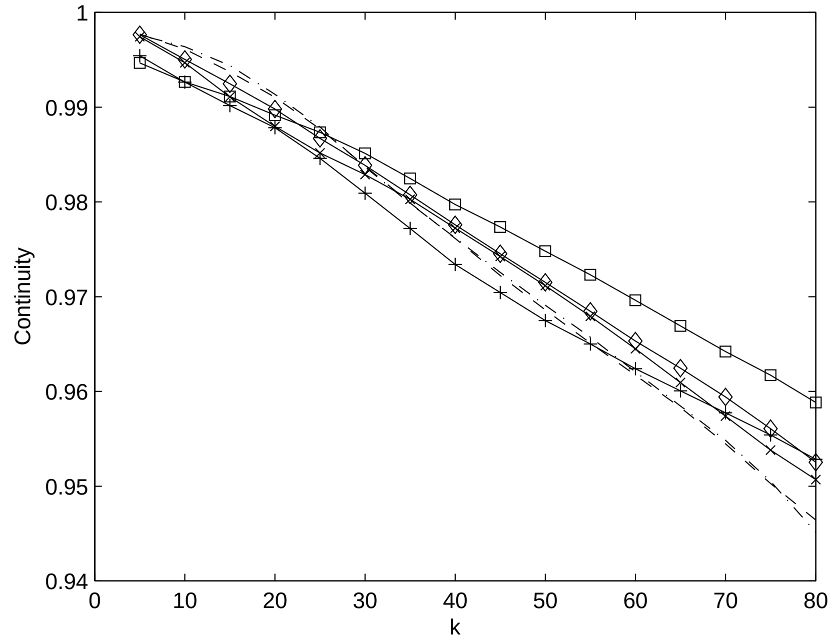


dz-dw representation

Chainlink: Performance measures

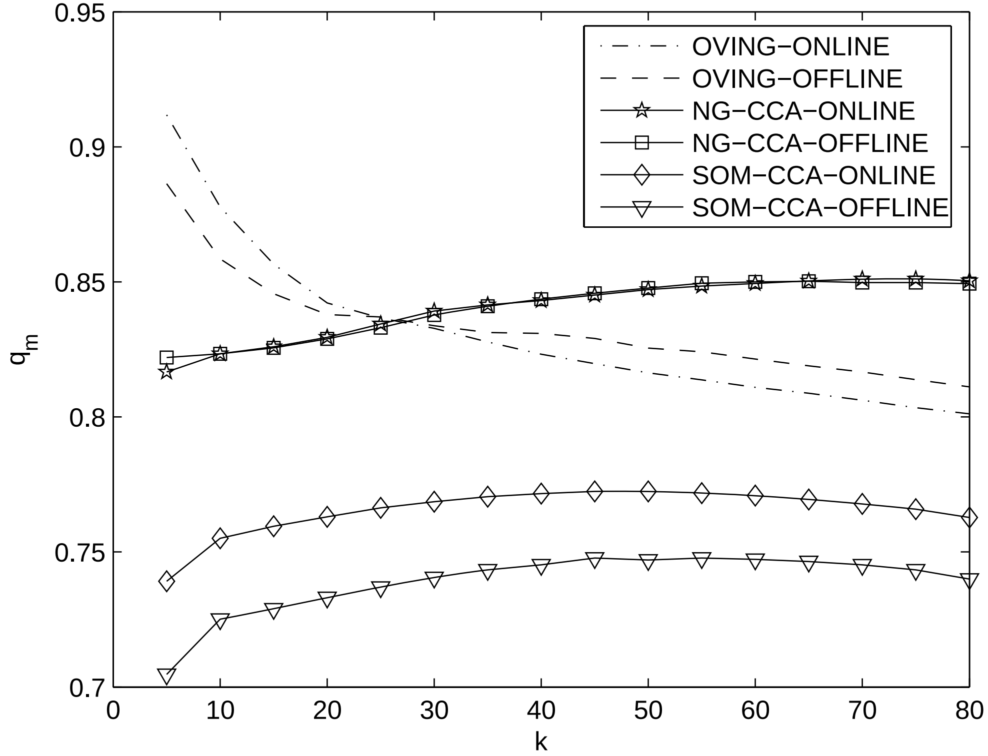


Trustworthiness as a function k



Continuity as a function of k

Chainlink: topology preservation



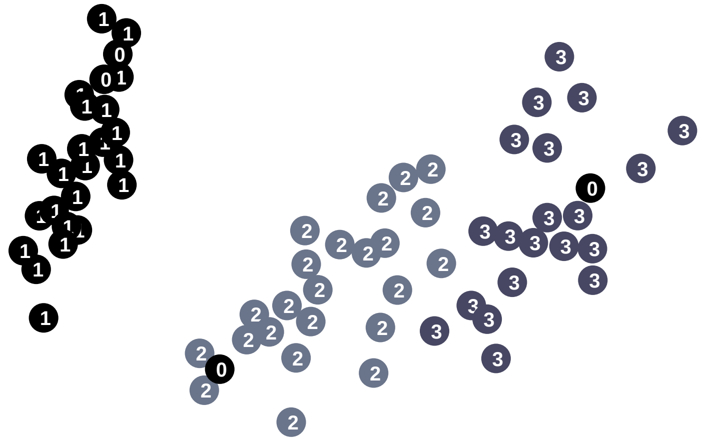
qm topology preservation measure as a function of k

Iris Data Set

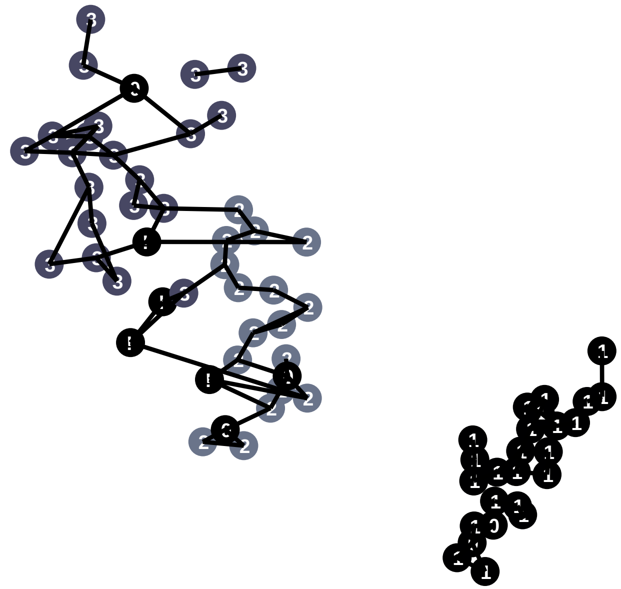
– 3 classes of Iris flowers: setosa, versicolor and virginica
– 150 examples with 4 attributes

Iris Flower

Iris Results

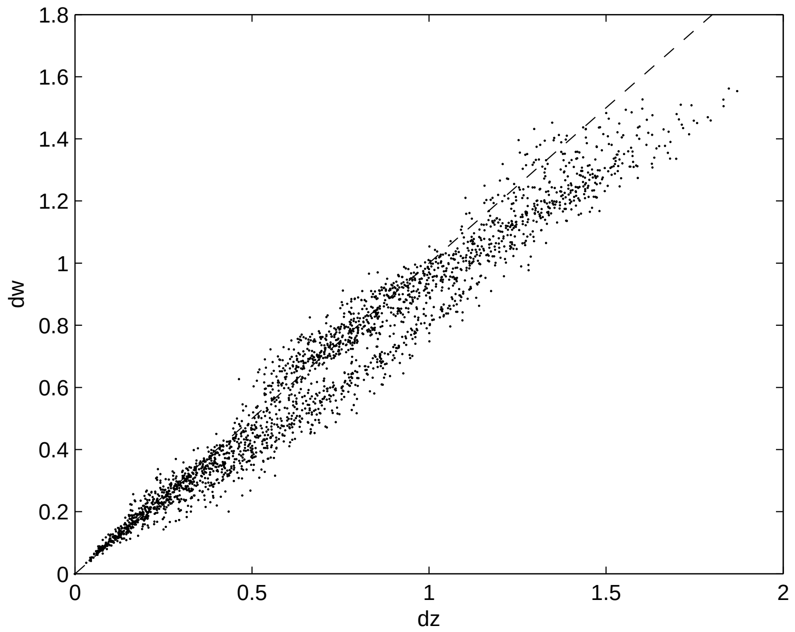


Quantized data projection with OVING online

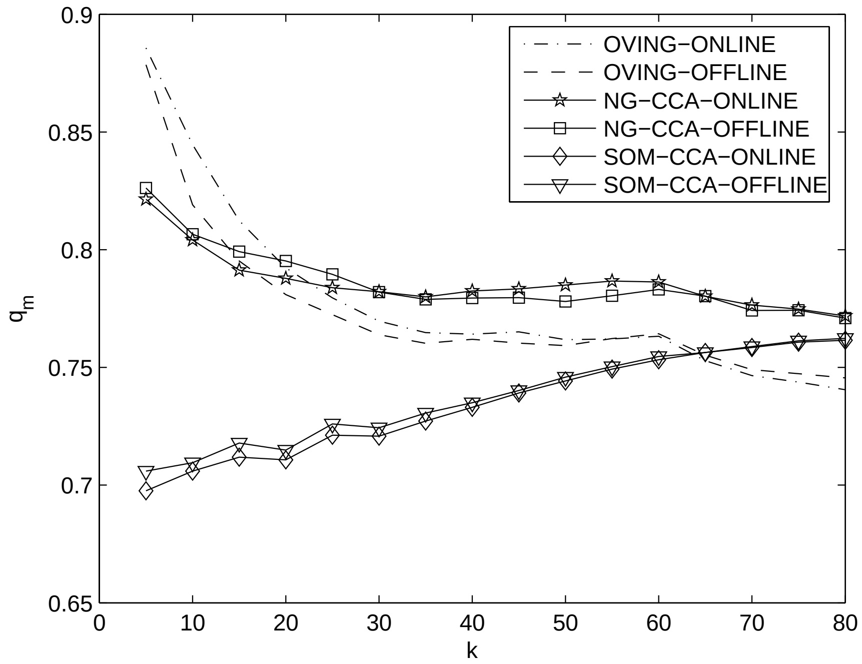


Iris data projection with CHR connections

Iris Data: Topology Preservation



dz-dw representation

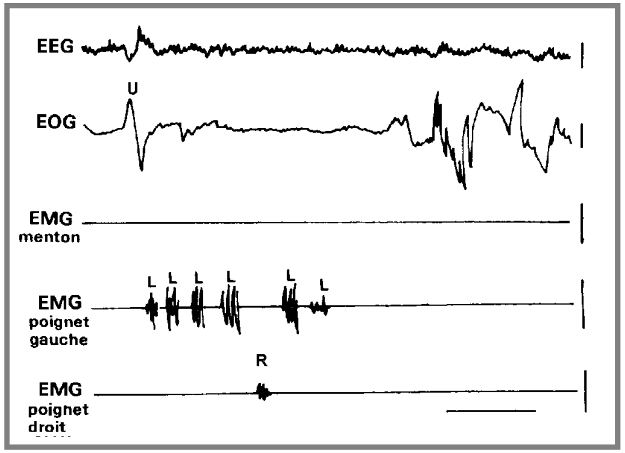


Topology preservation
measure qm

Sleep data set

- Real-world problem
- 6463 examples with 6 attributes
- Type of signals:
  - Rapid eye movements (REM)
  - Theta waves (TW)
  - Slow Delta (SD)
  - Sleep Spindles (SZ)
  - Muscular tone (MT)
  - Wakefulness detector

- 5 classes
  - Non-REM I (red)
  - Non-REM II (blue)
  - Non-REM III&IV (green)
  - REM (yellow)
  - Wakefulness (magenta)



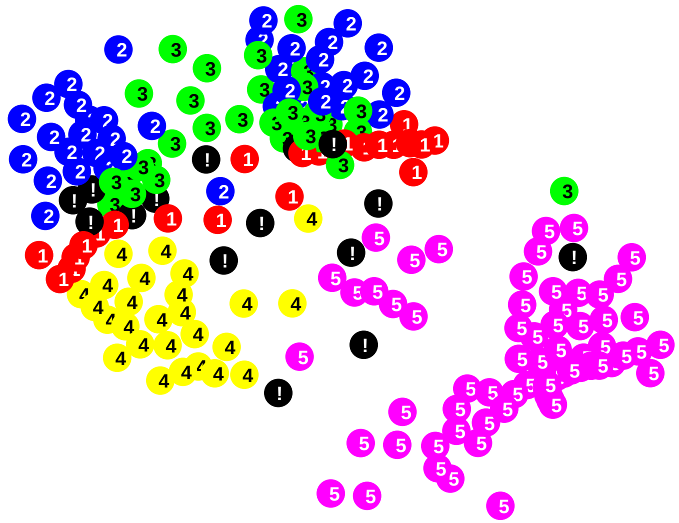
Polysomnographic recording

Pattern	Non-REM I	Non-REM II	Non-REM III&IV	REM	Wakefulness
REM	A	A	A	P	P
TW	P	X	X	P	X
SD	A	A	P	A	A
SZ	A	P	X	A	A
MT	X	X	X	A	P

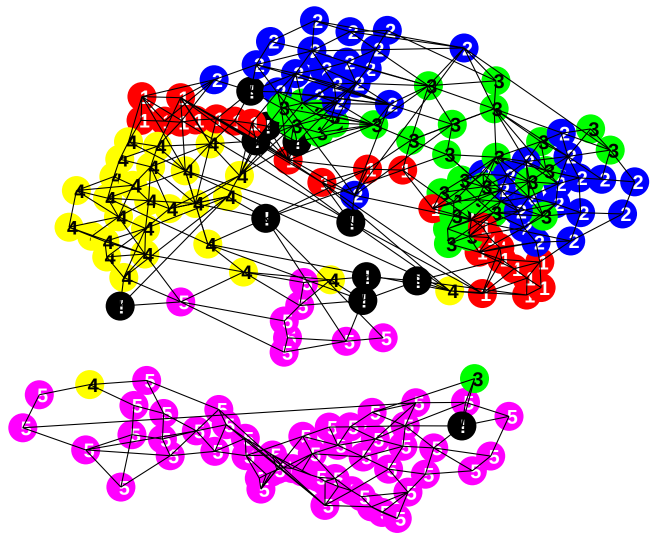
A: Absence

P: Presence

Sleep results

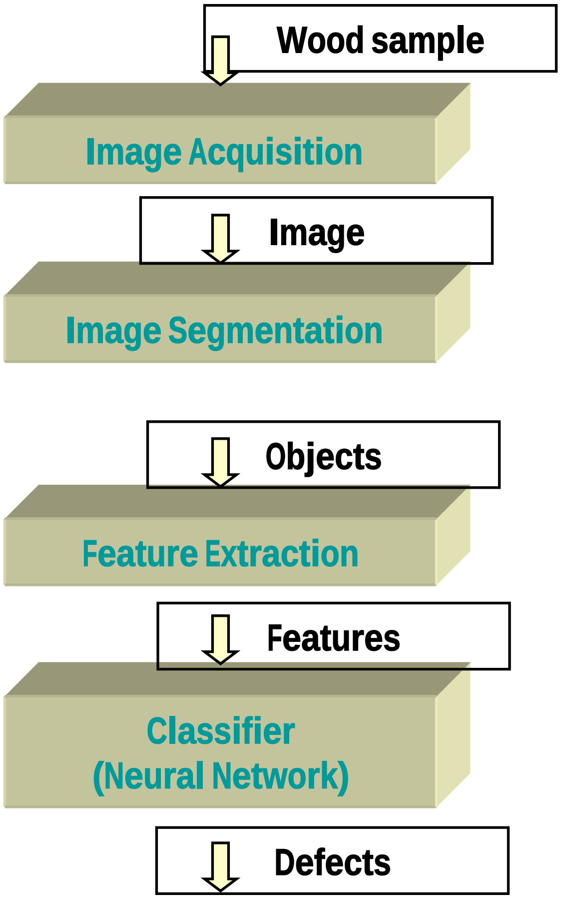


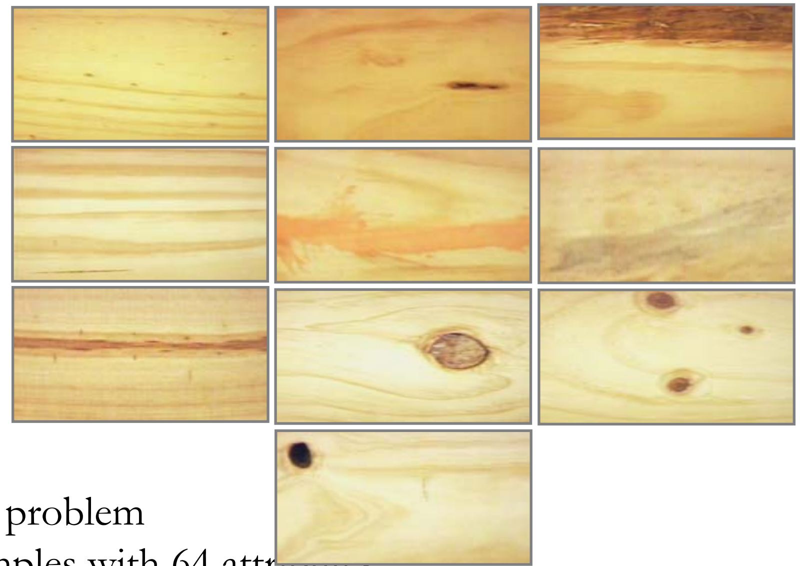
Quantized data projection
with OVING



Data projection with
CHR connections

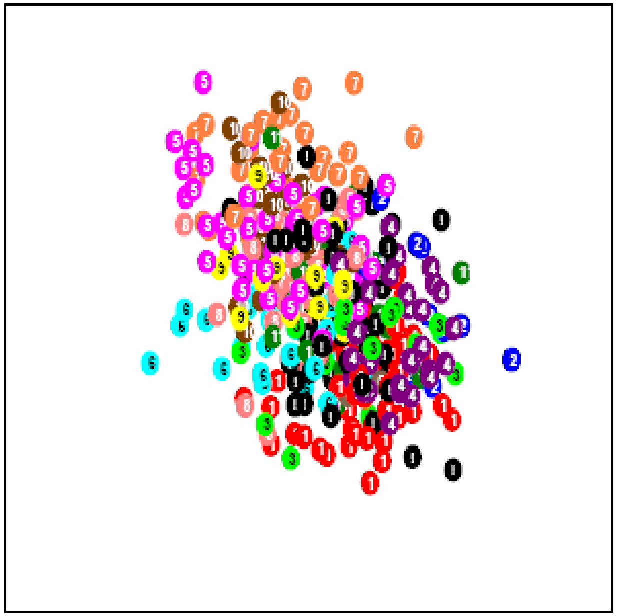
Wood data set



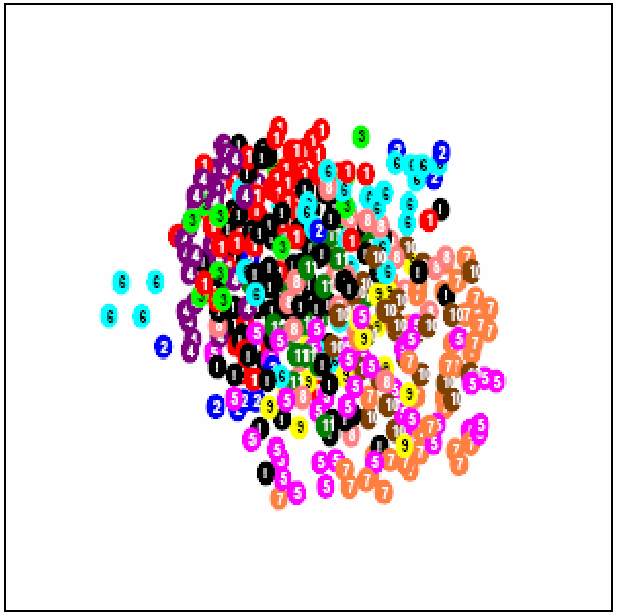


- Real-world problem
- 16800 examples with 64 attributes
- 11 classes:
  - Clearwood (red)
  - Birds eye (green)
  - Pocket (blue)
  - Wane (yellow)
  - Split (magenta)
  - Blue stain (cyan)
  - Stain (brown)
  - Pith (dark green)
  - Dead knot (navy blue)
  - Live knot (gray)
  - Hole (black)

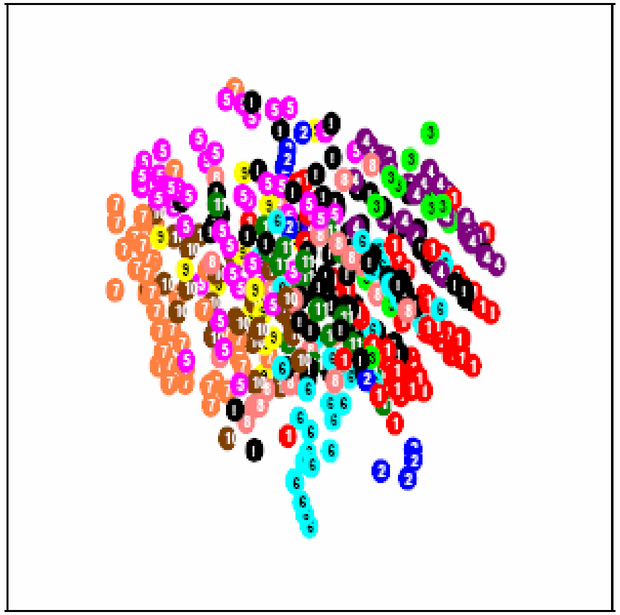
Wood results



DIPOL-SOM online



NG/NLM



OVI-NG

Algorithm	q_m
OVI-NG	$\mathbf{0.3467 \pm 0.0114}$
OVI-NG-2	0.2977 ± 0.0019
Dipol-SOM online	0.2219 ± 0.0187
Dipol-SOM offline	0.2179 ± 0.0227
SOM/NLM	0.3107 ± 0.0219
NG/NLM	0.2054 ± 0.0267
TOPNG	0.2483 ± 0.0051

Conclusions

- OVI-NG method provides an output representation to the NG that is useful for data projection and visualization
- It is an online method that simultaneously adjusts both the codebook vectors and the codebook positions
- It is computationally efficient with $O(N \log N)$ complexity
- OVI-NG obtained the best results in terms of the topology preservation measurement, qm, and the trustworthiness and continuity measures for local neighborhoods.

Future Work

- OVI-NG could be extended to map all data points
- Curvilinear or graph distances could be considered in order to unfold strongly nonlinear maps
- Other topology preservation measurement could be used such as the Topographic Product

OPTIONAL PART

Applications of Kohonen’s Map

- Text Retrieval in Very Large Document Collections
- Content-Based Image Retrieval
- Dimensionality reduction
- Visualization of high-dimensional data in 2D or 3D maps

Text Retrieval Problem

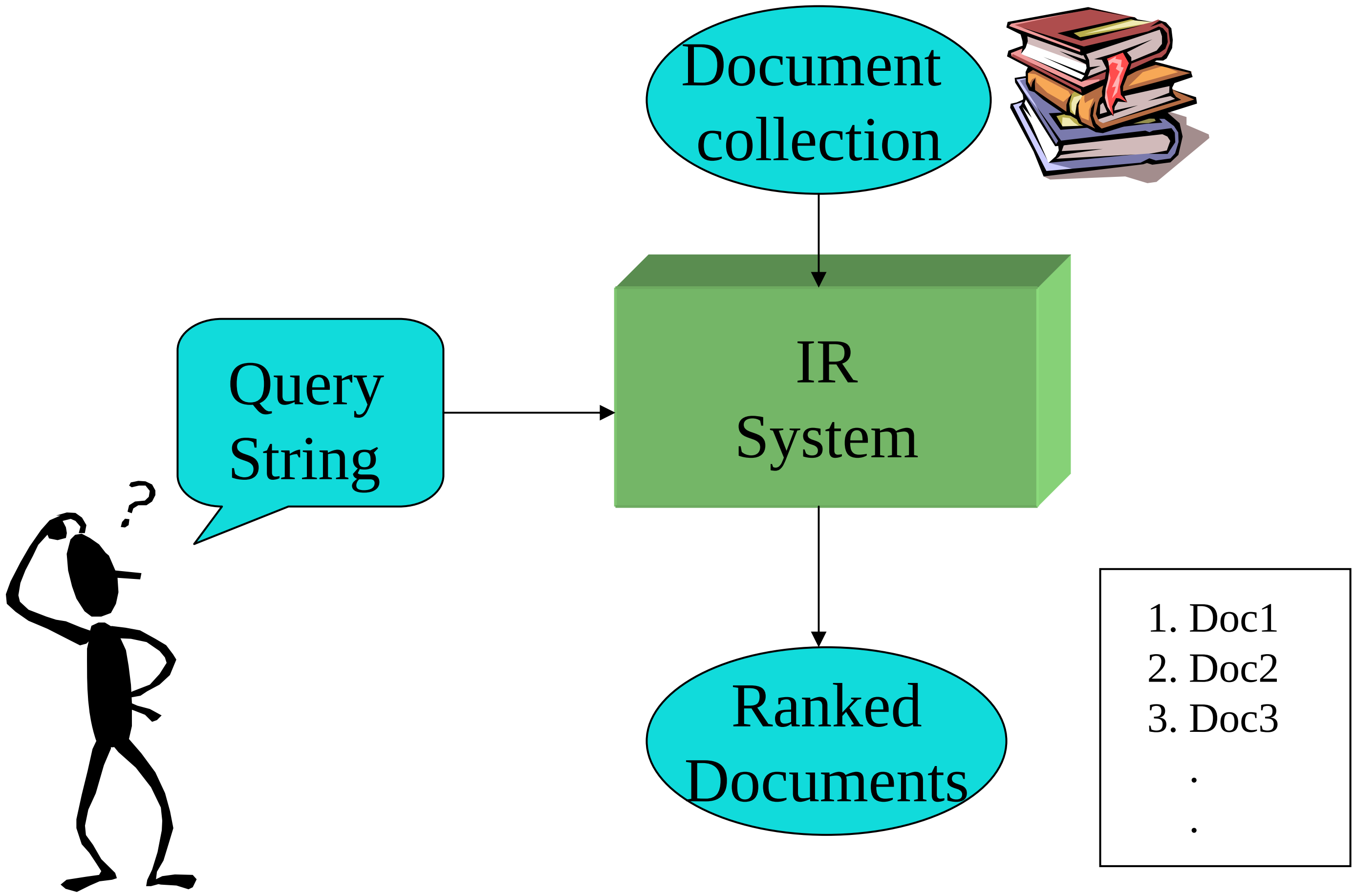
GIVEN:

- A large collection of textual documents in natural language
- A user query in the form of a textual string

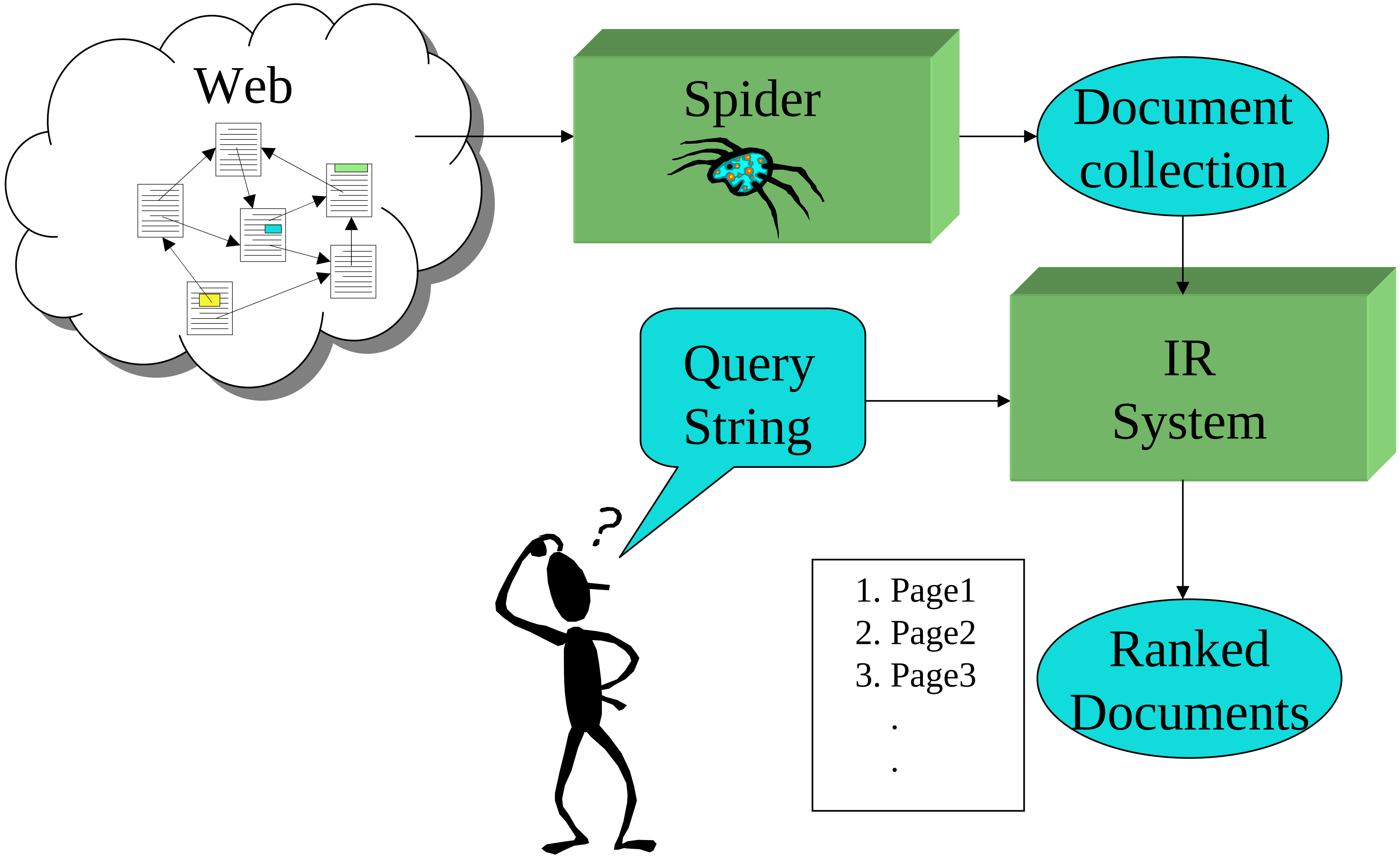
FIND:

- A ranked set of documents that are relevant to the query

Information Retrieval Architecture



Example: A Web Search System



Basic IR Approach: Keyword Query

- The assumption is that the keyword appears in the document collection.
- The keyword can be a simple word or a complex sentence.
- Terms can be combined by Boolean logic (AND, OR, etc.)

Limitations of Keyword Searching

- The user needs to know the domain to select an appropriate keyword
- Synonyms are not taking into account
  - “United States” vs. “USA”
- The context is lost
  - “Apple” (company vs. fruit)
  - “bit” (unit of data vs. something held with the teeth)

Vector Space Model (VSM)

- A popular method for document encoding
- The documents and queries are represented by vectors in which each component corresponds to one word
- The value of a component is a function of the frequency of occurrence of the word in the document and of its importance
- The relevance of a document to a query is measured using the similarity of the corresponding vectors

Representing a set of documents

- Using VSM, a set of Q documents containing R words is represented by a matrix M of dimension RxQ
- Each matrix element corresponds to the “weight” of a term in the set of documents

$$M = (m_{ij}) = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,Q} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,Q} \\ \vdots & \vdots & \ddots & \vdots \\ m_{R,1} & m_{R,2} & \cdots & m_{R,Q} \end{bmatrix}$$

Weights m_{ij}

- Let n_i/Q be the frequency of the ith-word in the collection of documents
- IDF: Inverse Document Frequency, where words that occur in few documents obtain larger weights
- The weight m_{ij} of the ith-word in the jth-document is

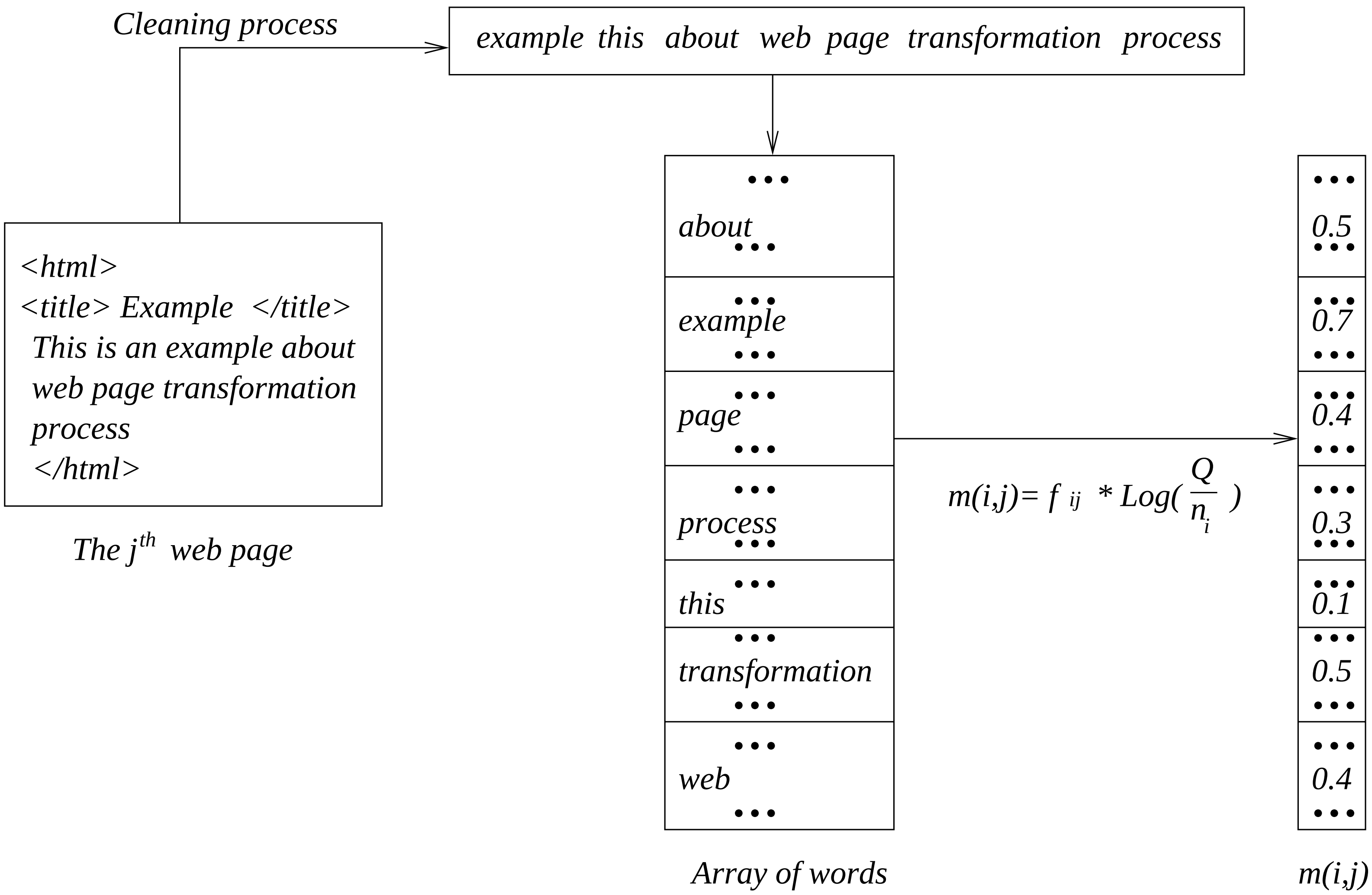
  where f_{ij} es the number of ocurrences of the ith-word in the jth-document

$$TF * IDF = f_{ij} * \log(Q / n_i)$$

Data Cleaning and Preprocessing

- Non-textual information is removed
- Articles and common words like pronouns, prepositions, conjunctions, etc. are deleted
- A word stemming process is applied to remove suffixes: e.g., `writing' to `write'; `are', `did', `were' to `be'
- A thesaurus may be used to take into account synonymous

Example of Vector Space Model



Web page Array of words Vector

Web page Array of words Vector

Similarity Measure

- The similarity between two vectors representing documents, or a search string and a document can be measured as

- Co $sim(d_i, d_j) = \frac{\vec{d}_i \cdot \vec{d}_j}{\left|\vec{d}_i\right|\left|\vec{d}_j\right|} = \frac{\sum_{k=1}^{R} m_{i,k} m_{j,k}}{\sqrt{\sum_{k=1}^{R} m_{i,k}^2}\sqrt{\sum_{k=1}^{R} m_{j,k}^2}}$

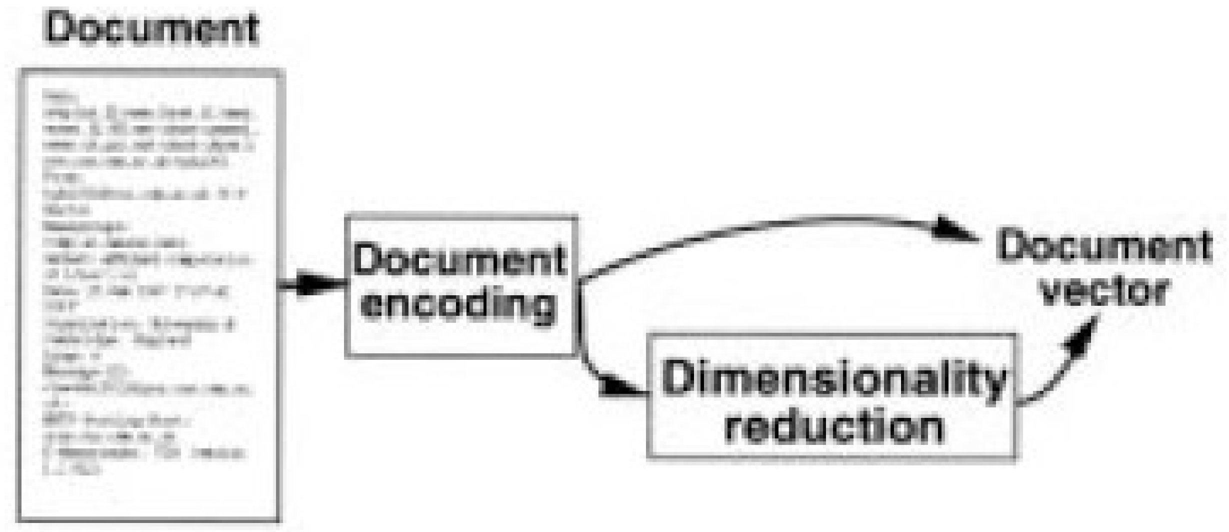
Limitations of VSM

- The contextual information is missed
  - VSM assume that each word is independent, does not take into account semantic relationships
  - The syntactic information is also missed (e.g. phrase structure, word order, proximity information).
- The dimensionality of the vector could be huge for large document collections

WEBSOM

- Automatic organization of a massive document collection according to textual similarities
- Graphical display map provides an overview of the collection
- Similar documents are found near each other
- Interactive browsing
- Scalable to very large document collections

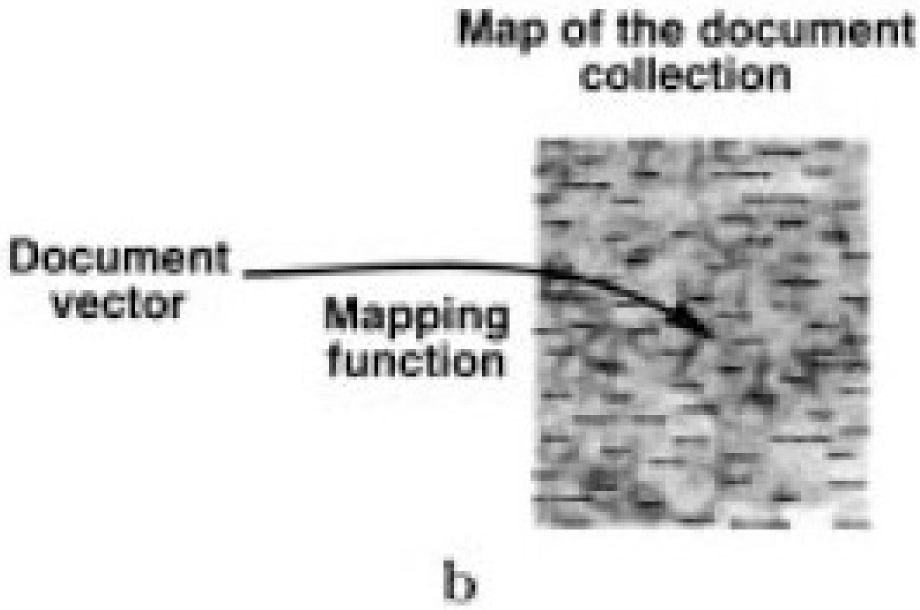
WEBSOM System (I)



a

- First the document is encoded into a numerical vector
- The dimensionality of the vector can be reduced

WEBSOM System (II)



- **Secondly, the numerical representation of the document is mapped onto a display of the document collection by using SOM**

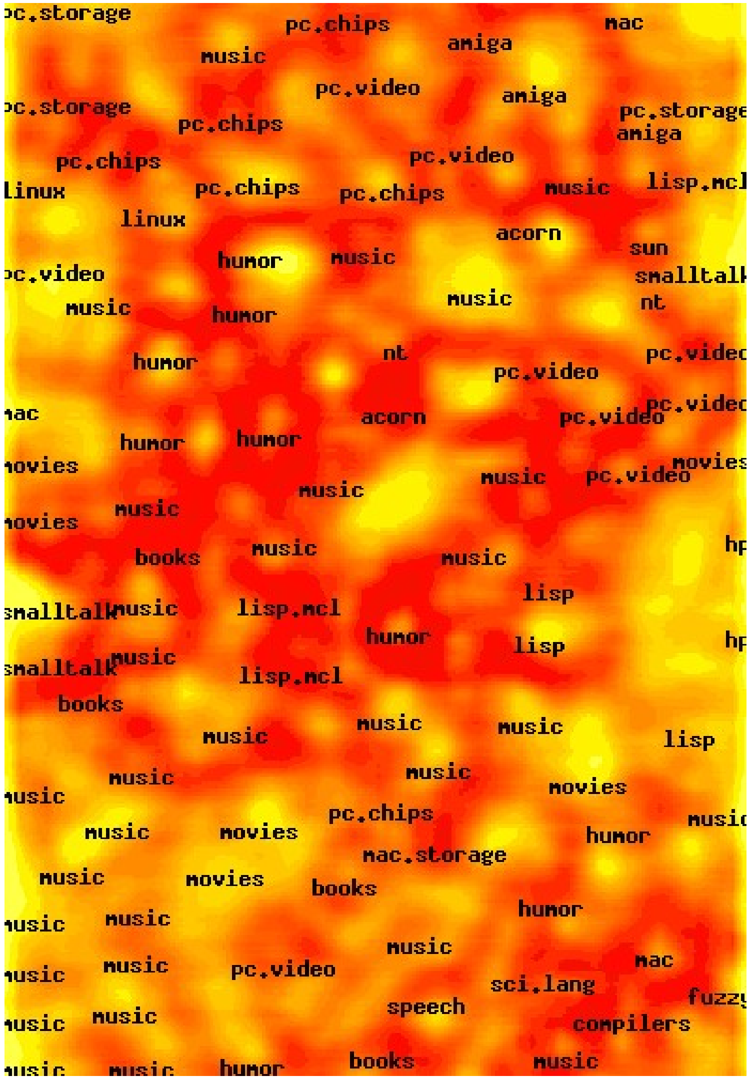
Application: A Million Documents from 85 Usenet Newsgroup

- amiga - comp.sys.amiga.hardware
- booksEach white20 - rec.arts.books
- cdrom - comp.publish.cdrom.hardware
- compilers - comp.compilers
- fuzzy - comp.ai.fuzzy
- genetic - comp.ai.genetic
- hp - comp.sys.hp.hardware
- acorn - comp.sys.acorn.hardware
- humor - rec.humor
- lang.eiffel - comp.lang.eiffel
- lang.ml - comp.lang.ml
- linux - comp.os.linux.hardware
- lisp - comp.lang.lisp
- lisp.mcl - comp.lang.lisp.mcl
- mac.storage - comp.sys.mac.hardware.storage
- movies - movies
- mac - comp.sys.mac.hardware.misc
- music - music
- nt - comp.os.ms-windows.nt.setup.hardware
- pc.cdrom - comp.sys.ibm.pc.hardware.cd-rom
- pc.chips - comp.sys.ibm.pc.hardware.chips
- pc.comm - comp.sys.ibm.pc.hardware.comm
- pc.storage - comp.sys.ibm.pc.hardware.storage
- pc.video - comp.sys.ibm.pc.hardware.video
- philosophy - philosophy
- plant - bionet.biology.plant
- prolog - comp.lang.prolog
- sci.lang - sci.lang

Newsgroups Application

- About 1,100,000 million documents written in English
- Average length of each text was 218 words
- Size of the SOM was 104,040 prototypes
- Whole vocabulary contained 1,127,184 unique words, which was reduced to 63,773 words after stemming and filtering
- Final dimensionality was reduced to 315 by using the random projection method

WEBSOM Map - Million Documents

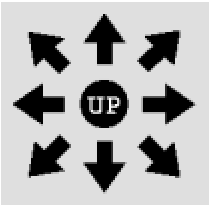


Browsing: When clicking on the map image with the mouse, a zoomed view of the part of the map will appear.

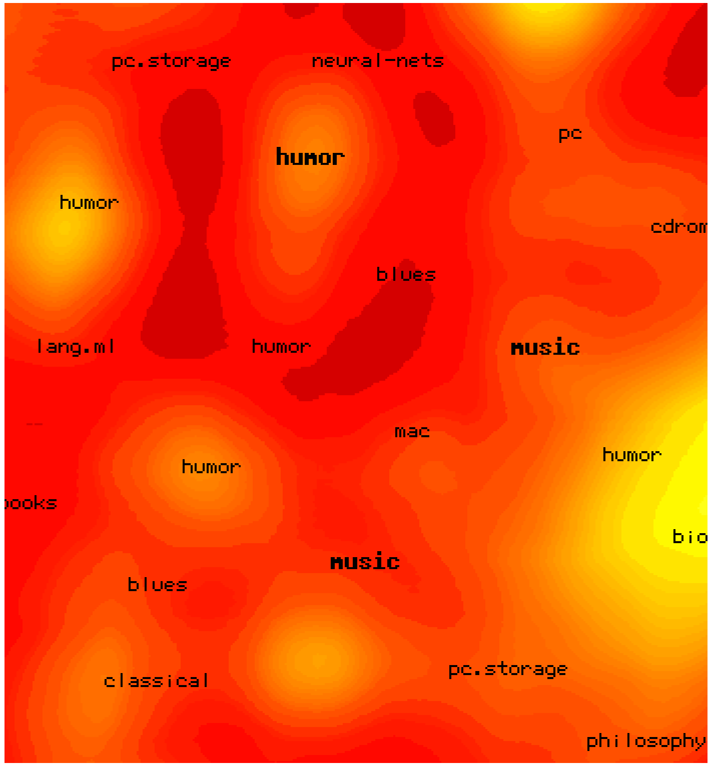
Labels: Give a general idea of the document collections

Coloring: Yellow areas are clusters and red areas empty spaces.

WEBSOM Zoomed Map

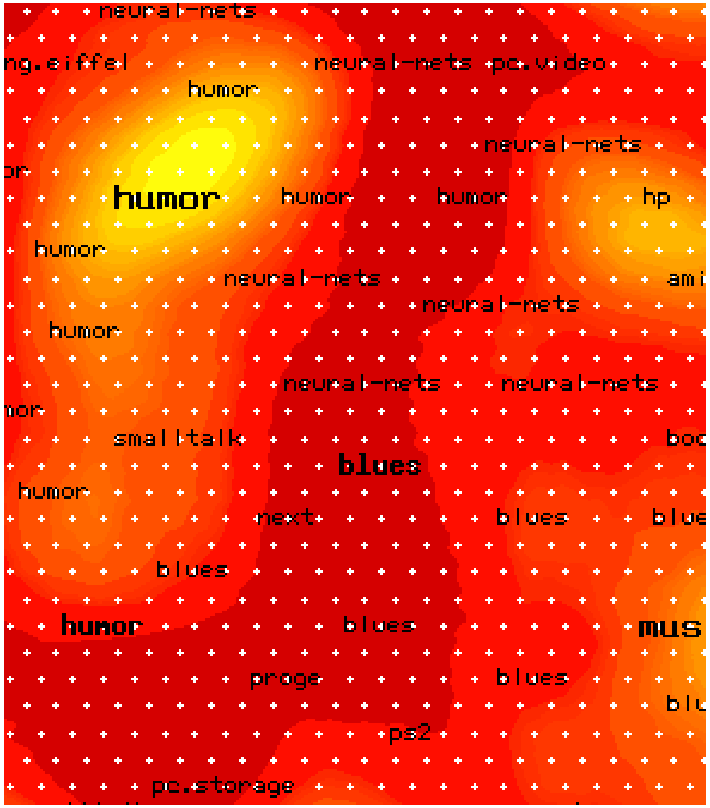


Click arrows to move to neighboring areas on the map, and to move up to the overall view.



- **blues - rec.music.bluenote**
- **books - rec.arts.books**
- **cdrom - comp.publish.cdrom.hardware**
- **classical - rec.music.classical**
- **humor - rec.humor**
- **lang.ml - comp.lang.ml**
- **mac - comp.sys.mac.hardware.misc**
- **music - music**
- **neural-nets - neural-nets**
- **pc - comp.sys.ibm.pc.hardware.misc**
- **pc.storage - comp.sys.ibm.pc.hardware.storage**
- **philosophy - comp.ai.philosophy**

WEBSOM Zoomed Map (II)



Click any white dot to enter the node, and see the contents of the individual map unit

Websom node cc 183

- **rec.music.bluenote**
  - **Re: DOLPHY/BLUENOTE?????** **Alan Saul, 6 Jun 1995, Lines: 35.**
  - **Rude boys, leave your guns at home!** **Jeremy Mushlin, 6 Jun 1995, Lines: 33.**
  - **Newbie seeks opinions** **RJMATTER@delphi.com, 7 Jun 1995, Lines: 51.**
  - **Re: George Barrow?** **Dennis Whitling, 7 Jun 1995, Lines: 10.**
  - **Re: Re: George Barrow?** **"Ashot W Bakunts", Wed, 07 Jun 95, Lines: 21.**
  - **Re: George Barrow?** **Laurie Sonnenfeld, 7 Jun 1995, Lines: 22.**
  - **Re: Time: How to feel it?** **Marc Sabatella, 7 Jun 1995, Lines: 30.**
  - **Re: Mingus/The Black Saint and the Sinner Lady** **Todd Adamson, 11 Jun 1995, Lines: 18.**
  - **Re: Mingus/The Black Saint and the Sinner Lady** **Gerald D. Wright, Wed, 14 Jun 1995, Lines: 22.**
  - **Re: Music to bury me by** **TornCot, 14 Jun 1995, Lines: 11.**
  - **Re: Joe Henderson Double Rainbow Tour** **Thomas F Brown, 23 Jun 1995, Lines: 10.**
  - **Re: Coltrane reissues on Impulse** **Kyung-Goo Doh, 03 Jul 1995, Lines: 23.**

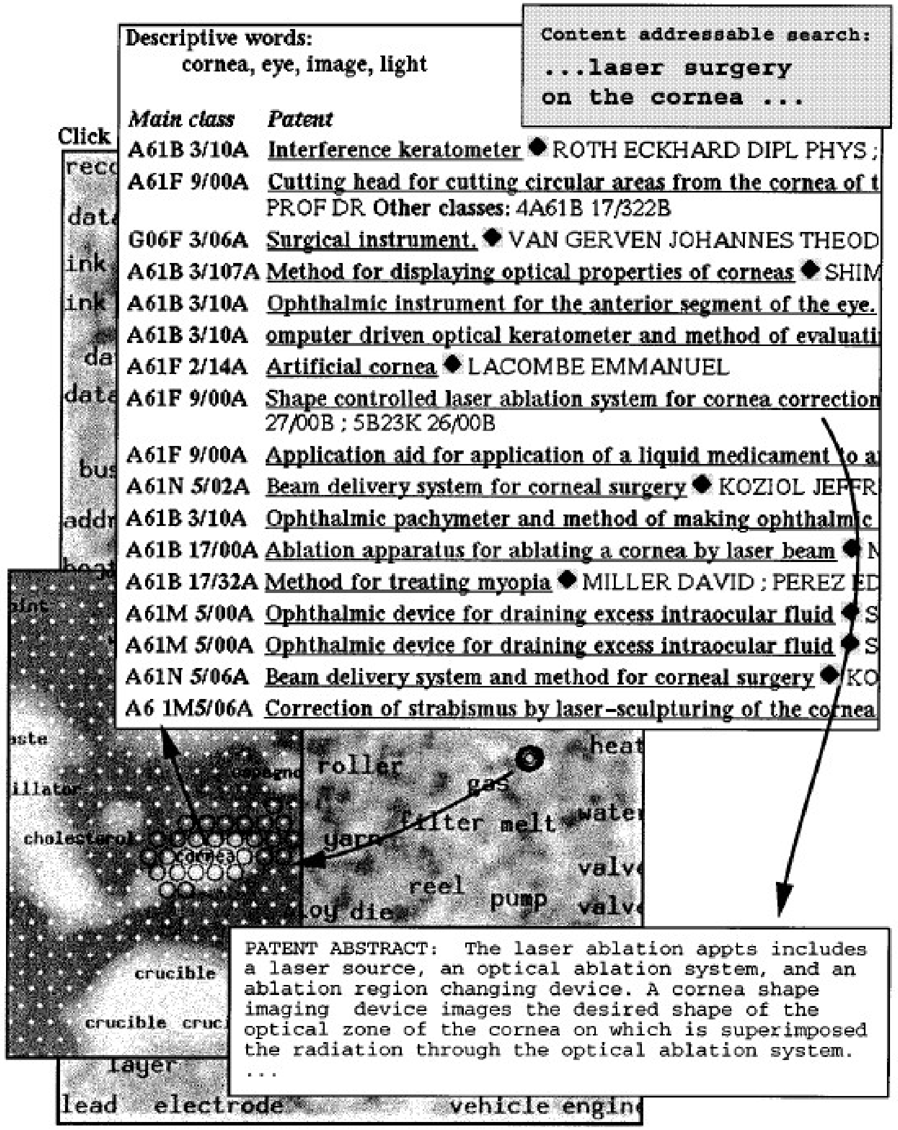
Application: Corpus of US, Japanese and European Patents

- About 7,000,000 million patent abstracts written in English
- Average length of each text was 132 words
- Size of the SOM was 1,002,240 prototypes
- Whole vocabulary contained 733,179 words, which was reduced to 43,222 words after stemming and filtering
- Final dimensionality was reduced to 500 by using the random projection method

Content Addressable Search

- The user can type a query in the form of a short document
- The query is represented by a vector
- The resulting vector is compared with the prototypes of all map units
- The best matching points are marked with circles
- These locations provide good starting points for browsing

Content Addressable Search (II)



Query: Laser surgery on the cornea

Best Matching: Marked with circles

Zooming: Reveals a cluster of map units containing patents about the cornea of the eye

PicSOM

- Content-based image retrieval with SOM in large databases
- Creates a 2D map, where similar images are found near each other
- Iterative process utilizing the relevance feedback approach
- The initial query begins with an initial set of images selected from the database

Conclusions

- **Unsupervised neural networks are usually applied for clustering and visualization of high-dimensional data**
- **Patterns in high-dimensional feature space can be projected onto 2D or 3D maps**
- **Neural networks techniques can help to improve the traditional information retrieval methods**
- **Neural maps could be useful for browsing and exploratory analysis**

Conclusions (I)

- **Unsupervised neural networks are usually applied for clustering and visualization of high-dimensional data**
- **Patterns in high-dimensional feature space can be projected onto 2D or 3D maps**
- **Ideally the projections should be distance and topology preserving to avoid false interpretations**
- **There exist quality measures of distance and topology preservation**

References

- V. Cherkassky, F. Mulier, Learning from Data, Wiley Interscience, 1998

- http://websom.hut.fi/websom/

- http://www.cis.hut.fi/picsom/

- T. Kohonen et al., Self-organization of a Massive Document Collection, IEEE Transactions on Neural Networks, Vol. 11, pp. 574-585, May 2000

References (II)

- **R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, Addison-Wesley, 1999**