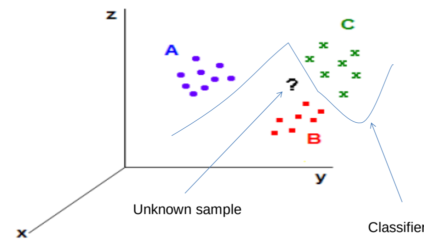
Classification of continuous multi-way data via dissimilarity representation

Author: Diana Porro Muñoz

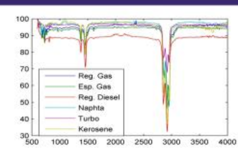
Supervisors: Robert P.W. Duin (PRLab, TUDelft)
Isneri Talavera Bustamante (CENATAV)

Introduction. Supervised Classification

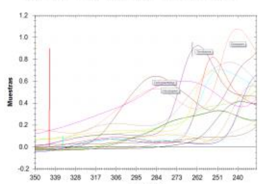


Spectral data

FT-IR espectrómetro



UV-VIS espectrómetro



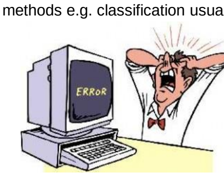
Representation

1	var1	var2	...	...	...	...	...	...	...	var m
2	var1	var2	...	...	...	...	...	...	...	var m
. n	var1	var2	...	...	...	...	...	...	...	var m

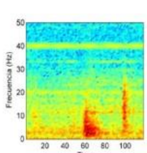
Data set (n<< m)

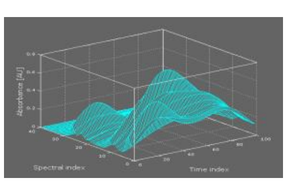
Curse of dimensionality!!!

Traditional methods e.g. classification usually fail to learn well!!

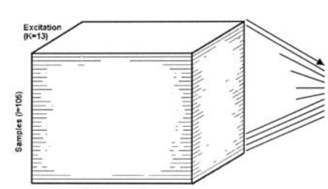


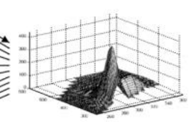
Examples of multi-dimensional data





Representation of continuous multi-way objects



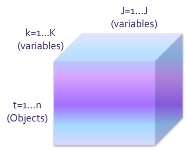


$\underline{\mathbf{X}}\ (I \times J \times K)$

Multi-way data analysis

- So, an intuitive way to represent this relationship for all the objects would be in a multi-way array

$$Y \in \mathbb{R}^{C_1 \times C_2 \times \cdots \times C_z}$$



$$n \ll J \times K$$

$$C_1 \ll C_2 \times \cdots \times C_z$$

Existing methods

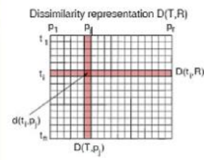
- Regression
- Exploratory analysis

Supervised classification????

- Unfolding (1D) + classifiers
- Decomposition methods + classifiers
- NSIMCA

Dissimilarity Representation (DR)

- It is based on the role that (dis)similarities play in a class composition, where objects from the same class should be similar and objects from different classes should be different (compactness property). Hence, it should be easier for the classifiers to discriminate between them.
- It proposes to train classifiers in the space of the proximities between objects, instead of the traditional feature.



$$R(p_1, p_2, \cdots, p_r) \Longrightarrow \{R \subseteq T, R \equiv T, any\ other\}$$

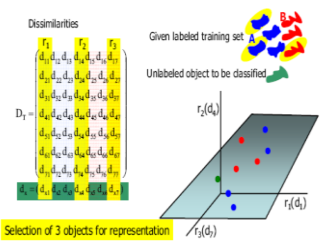
$$\emptyset(\cdot, R): \mathbb{R}^m \rightarrow \mathbb{R}^r$$

Objects are represented in this space by the column vectors of the dissimilarity matrix.

$$\emptyset(t_i, R): [d(t_i, p_1), d(t_i, p_2), \cdots, d(t_i, p_r)]$$

Dissimilarity Representation (DR)

Relationship between objects is used for the classification. With a suitable measure, the compactness property should be fulfilled, thus it should be easier to discriminate between classes.



Some advantages of DR for our problem

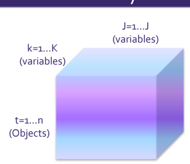
✓ Dimensionality is reduced to the number of objects
✓ Any knowledge or information about the problem background can be included into the dissimilarity measure.
✓ Any traditional classifier can work on the Dissimilarity Space!!!

+

DR can be generated from many initial representations of the objects eg. numerical vectors, graphs, as long as a "suitable" measure is used.

Porro-Muñoz D, Talavera I, Duin R.P.W, Hernández N. The representation of chemical spectral data for classification. Proceedings CIARP 2009, vol. 5856, LNCS, Springer: Guadalajara, Mexico, 2009, 513–520.
Porro-Muñoz, D., Talavera, I., Duin, R.P.W., Hernández, N., Orozco-Alzate, M.: Dissimilarity representation on functional spectral data for classification. Journal of Chemometrics 25 (2011) 476–486.

DR to classify three-way data?



?

$$d_{ir} = D(t_i, p_j)$$

$$Y \in \mathbb{R}^{C_1 \times C_2 \times \cdots \times C_z}$$

$$\emptyset(\cdot, R): \mathbb{R}^{C_1 \times C_2 \times \cdots \times C_{z-1}} \rightarrow \mathbb{R}^r$$

$$\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1r} \\ d_{21} & d_{22} & \dots & d_{2r} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nr} \end{bmatrix}$$

2D Shape measure for three-way spectral data

$$D^1_{a,b} = \left(\sum_{k=1}^{K} \left(\sum_{j=1}^{J} \left(y^{\sigma}_{a,j,k} - y^{\sigma}_{b,j,k} \right)^2 \right)^{p/2} \right)^{1/p}, y^{\sigma}_{i,j,\cdot} = \frac{d}{d_j} G(j,\sigma) * y_{i,j,\cdot}$$

1D analysis
1st direction

$$D^2_{a,b} = \left(\sum_{j=1}^{J} \left(\sum_{k=1}^{K} \left(y^{\sigma}_{a,j,k} - y^{\sigma}_{b,j,k} \right)^2 \right)^{p/2} \right)^{1/p}, y^{\sigma}_{i,\cdot,k} = \frac{d}{d_k} G(k,\sigma) * y_{i,\cdot,k}$$

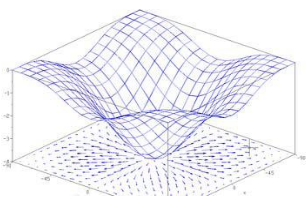
1D analysis
2nd direction

$$D = \frac{1}{w_1} D^1 + \frac{1}{w_2} D^2$$

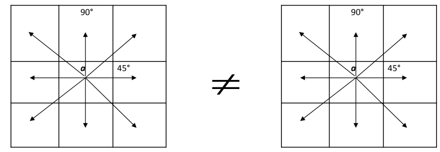
Porro-Muñoz, D., Duin, R.P.W., Orozco-alzate, M., Talavera, I., Londoño Bonilla, J.M. The dissimilarity representation as a tool for three-way data classification: a 2D measure. In: Proceedings of S+SSPR2010. Volume 6218, LNCS, Springer: C¸ e¸sme, Turkey, 2010, 569–578.
Porro-Muñoz D., Duin R.P.W., Talavera I., Orozco-Alzate M. Classification of three-way data by the dissimilarity representation. Signal Processing 2011; 91(11): 2520–2529.

Gradient Measure

- Vector field: We are taking into account how shape changes for each point in all directions.



Continuous Multi-way Shape Measure (CMS)



As derivatives are undefined for discrete functions, they need to be estimated somehow to be used on these data. A widely used method for approximating the derivative of a discrete function is the application of linear filters by convolution.

Definition 1. *Let $\underline{Y}$ be a n-way data set and let $\underline{Y_a}$, $\underline{Y_b}$ be two objects from this data set. The dissimilarity between $\underline{Y_a}$ and $\underline{Y_b}$ can be computed as:*

$$d_G(\underline{Y_a}, \underline{Y_b}) = \left\| \sum_{i=1}^{f} \underline{Y_a} * \underline{G_\sigma} * \underline{H_i} - \underline{Y_b} * \underline{G_\sigma} * \underline{H_i} \right\|_F \quad (1)$$

where $\|\cdot\|_F$ is the Frobenius norm for tensors [8], $\underline{G_\sigma}$ a Gaussian convolution kernel to smooth the data first, $\underline{H_i}$ is a partial derivative kernel and f is the amount of partial derivatives in the different directions in order to obtain the gradient.

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

$$m = \frac{\Delta f(x)}{\Delta x} = \frac{f(x+h) - f(x)}{(x+h) - (x)} = \frac{f(x+h) - f(x)}{h}.$$

$H_x^P =$

-1	0	1
-1	0	1
-1	0	1

$H_y^P =$

-1	-1	-1
0	0	0
1	1	1

$H_{md}^P =$

-1	-1	0
-1	0	1
0	1	1

$H_{sd}^P =$

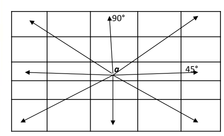
0	-1	-1
1	0	-1
1	1	0

The Prewitt and other defined gradient operators are based on approximating a partial derivative at a point ***a*** by computing the slope of the line that fits the previous and next point of ***a*** in the direction of the derivative.

Porro-Muñoz, D., Duin, R.P.W., Orozco-Alzate M., Talavera, I.: Continuous Multi-way Shape Measure for Dissimilarity Representation. In: Proceedings of the 17th Iberoamerican Congress on Pattern Recognition CIARP 2012. Volume 7441 of LNCS, 430–437.

Gradient Polynomial-Based Kernel for the CMS Measure

We propose to approximate each partial derivative in point ***a*** as the derivative of the polynomial of degree ***t***, which is obtained by interpolating ***a*** and its ***t*** nearest points in the direction of the derivative.



$p_0 = (-2, y_0)$,

$p_1 = (-1, y_1)$, $p_2 = (0, y_2)$, $p_3 = (1, y_3)$ and $p_4 = (2, y_4)$

$$P(x) = ax^4 + bx^3 + cx^2 + dx + e$$
$$P'(x) = 4ax^3 + 3bx_2 + 2cx + d$$
$$P'(0) = d = \frac{y_0 - 8y_1 + 8y_3 - y_4}{12}$$

$H_x^L =$

1	-8	0	8	1
1	-8	0	8	1
1	-8	0	8	1
1	-8	0	8	1
1	-8	0	8	1

$H_y^L =$

1	1	1	1	1
-8	-8	-8	-8	-8
0	0	0	0	0
8	8	8	8	8
-1	-1	-1	-1	-1

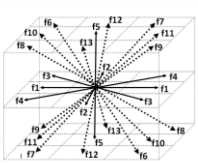
$H_{md}^L =$

0	0	0	1	1
0	0	-8	-8	1
0	8	0	-8	0
-1	8	8	0	0
-1	-1	0	0	0

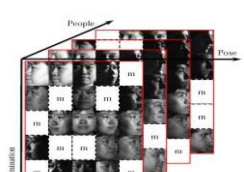
$H_{sd}^L =$

1	1	0	0	0
1	-8	-8	0	0
0	-8	0	8	0
0	0	8	8	-1
0	0	0	-1	-1

Four-way data



- For each point (13 directions -> 13functions)
- Example: Gas Chromatography/Fluorescence Objects x Time x Excitation x Emission, Sensor arrays

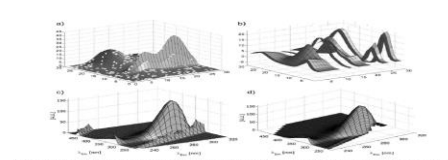


Other theoretical contributions

- Dimensionality reduction in both directions of three-way array.
- ✓ Reduces cost for computing dissimilarity matrix
- ✓ Noise is reduced thus classification results are improved

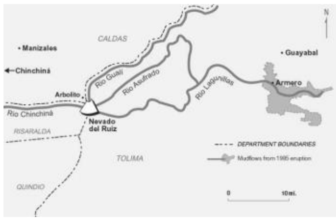
Optimizing Dissimilarities for the Classication of Three-way Chemical Spectral Data. D. Porro-Muñoz, Robert P. W. Duin, Isneri Talavera, Mauricio Orozco. Under major revision for Chemometrics and Intelligent Laboratory Systems

- Missing data analysis.

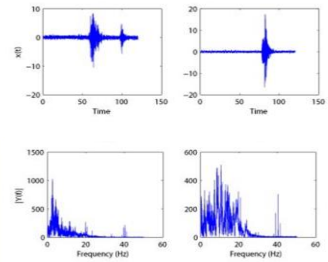


Classification of seismic volcanic data

The automatic classification of seismic volcanic signals is an essential task nowadays, with the goal of discovering the interaction between volcanic earthquakes and volcanic processes.



Traditional Representation of seismic volcanic data

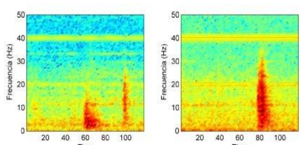


Does not take frequency changes along time into account

Does not take into consideration the time information

Time or frequency representations alone may not be optimal for seismic signal analysis, since spectral energy changes in time

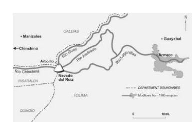
2D Representation



- Spectrograms: Time-frequency representation , showing frequency changes in time.
- So far, the spectrograms have just been averaged to obtain the spectral representation . The 2D object representations has not intensively been exploited as such in automatic classification systems.

Experiments with signals from Nevado del Ruiz Volcano

- Volcano: Nevado del Ruiz , Colombia
- Classes: Long Period, 235 events
  Volcano Tectonic earthquakes, 235 events
- Event: Time signals of 12032 points (120 sec.)
- 1D (spectral): 256-point Fast Fourier Transform (FFT).
- 2D (spectrogram): 256 short time Fourier transform
  Windows size: 256 points with 50\% of overlap.
- Data set: 470 x 129 x 93
- Fisher Linear classifier on DR.
- 10 times 10-fold cross-validation process.
- Prototypes were randomly chosen.



Volcano data set

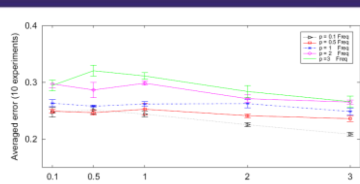


Fig 2. Average cross-validation error (with standard deviation) for the 2Dshape measure for different p-values on both directions for seismic volcanic data set.

Representations	Volcano
1D	30.2(0.4)
2D	20.9(0.2)

Results

Colon cancer (1000 x 60 bin x 90 scales): Normal and Tumor.
Quality process (323 x 15 excitation x 15 emission autofluorescence): Low and High
Age of Parma ham (67 x 15 excitation x 15 emission autofluorescence): 4 age ranges

	CMS						No shape	
Data	Prew.	Prew.(4d)	Sob.	Sob.(4d)	Polyn.	2Dshape	Frob	Yang
Colon cancer	11.0	11.5	11.2	12.0	**9.5**	12.7	13.3	13.3
Volcano	28.0	25.6	28.2	23.4	23.4	**20.9**	40.0	28.7
Enzyme	9.4	**5.7**	9.4	9.4	9.4	13.2	9.4	9.4
Parma ham	3.7	**2.4**	3.7	2.5	3.7	2.9	4.5	4.3
Carrot juice	**7.2**	**6.0**	**7.2**	**6.3**	**7.1**	**8.3**	9.8	10.7

Future Perspectives

- Towards the application in other research areas
- Clustering for multi-way data
- Dissimilarity Representation for Regression
- Dissimilarity Representation for non-continuous multi-way data

Thank you!!!!!