Clustering models for high dimensional, temporal, and dissimilarity data

Maurizio Vichi
Dip. Statistics Probability & Applied Statistics
University "La Sapienza" Rome
maurizio.vichi@uniroma1.it

1. Introduction

Often data present a multiway structure, and they can be arranged into a Three-way Data Set X, i.e.,a set **X** of $n \times K \times T$ values related to: K variables measured (observed, estimated) on n objects (individuals, products) at T occasions (assessors, times, locations, etc.). Let $X_1, X_2, ..., X_K$ be K quantitative variables observed on n units (objects) at T consecutive time points (Figure 1).

The observed data can be arranged into a three-way longitudinal data set

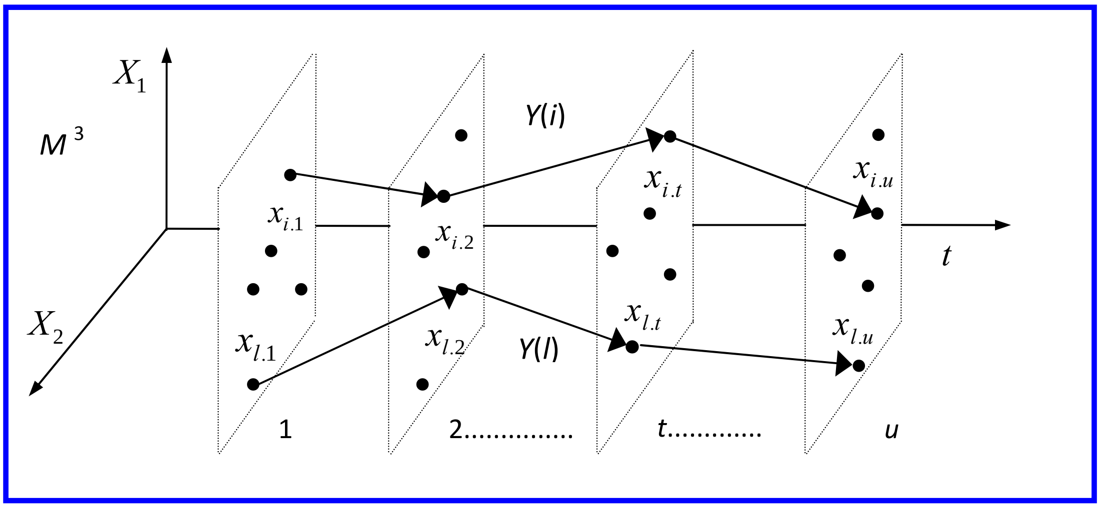
$$\mathbf{Y} \equiv \left[\mathbf{y}_{i \cdot t} = (x_{i1t}, x_{i2t}, ..., x_{ikt}, t) : i \in I, t = 1, ..., T\right]$$

where x_{ijt} is the value of the j-th variable collected on the i-th object at time t; $I=\{1,...,n\}$ $J=\{1,...,k\}$ and $U=\{1,...,T\}$ are the set of indices pertaining to objects,

variables and time points, respectively.

For each object i, $Y(i) = \{y_{i.t}$ t=1,...,$T\}$ describes a *time trajectory* of the i-th object

according to the k examined variables. The trajectory $Y(i)$ is geometrically represented by T-1 segments connecting T points $y_{i.t}$ of M^{k+1}. Two time trajectories in M^{3}



2. Trend, Acceleration and Velocities

The observed objects can be represented as points of a vectorial Let M^{k+1} be the metric space spanning the k variables and time. The problem to find a dissimilarity between trajectories is relevant. A distance between trajectories is defined as a function of distances between some characteristics of trajectories: VELOCITIES and ACCELERATIONS

***VELOCITY*:** Velocity of $Y(i)$ is defined as the rate of change of i-th object position in a fixed time interval and indicating the direction and versus of each segment of the trajectory $Y(i)$ for a given variable, i.e.:

$$v_{ijt,t+1} = \frac{x_{ijt+1} - x_{ijt}}{s_{t,t+1}} \tag{1}$$

Here $S_{t,t+1}$ is the interval from t to t+1.

In particular: $v_{ijt,t+1} > 0$ ($v_{ijt,t+1} < 0$) if object i, for the j-th variable, presents an increasing (decreasing) rate of change of its position in the time interval from t to t+1; $v_{ijt,t+1} = 0$ if the object i for the j-th variable, does not change position from t to t+1.

In M^2 velocity of each segment of the trajectory is the slope of the straight line passing through it. If velocity is negative (positive) the slope will be negative (positive) and the angle made by each segment of the trajectory with the positive direction of the t-axis will be obtuse (acute).

ACCELERATION measures the variation of velocity of $Y(i)$ in a fixed time interval.

For each time trajectory $Y(i)$, the acceleration of an object i in the interval from t to t+2 (Acceleration must be computed on two time intervals [t, t+1], [t+1, t+2]), denoted $s_{t,t+2}$, is, for the j-th variable:

$$a_{ijt,t+2} = \frac{v_{ijt+1,t+2} - v_{ijt,t+1}}{s_{t,t+2}}\,. \tag{2}$$

In particular: $a_{ijt,t+2}$>0 ($a_{ijt,t+2}$<0) if the object i, for the j-th variable, presents an increasing (decreasing) variation of velocity in the time interval from t to t+2; $a_{ijt,t+2}$=0 if object i, for j-th variable, does not change velocity from t to t+2.

Geometrically, acceleration of each pair of segments of trajectory represents their convexity or concavity. If acceleration is positive (negative) the trajectory of the two segments is convex (concave).

Defined velocity and acceleration, we are now in position to evaluate differences between trends in a time point t, velocities and accelerations in a time interval. Let us first consider:

TRENDS Distance

$$ {}_1\delta(i,l) = \pi_1 \sum_{t=1}^{T} \left\| \mathbf{X}_{i.t} - \mathbf{X}_{l.t} \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{X}_{..t}}} = \pi_1 \sum_{t=1}^{T} tr\left[\left(\mathbf{X}_{i.t} - \mathbf{X}_{l.t} \right)' \boldsymbol{\Sigma}^{-1}_{\mathbf{X}_{..t}} \left(\mathbf{X}_{i.t} - \mathbf{X}_{l.t} \right) \right] \qquad (3) $$

where π_1 is a suitable weight to normalize distances and $\boldsymbol{\Sigma}_{\mathbf{X}_{..t}}$ is the dispersion matrix of $\mathbf{X}_{..t}$. differences between trend intensities, in a time point t, of objects i and l are evaluated according to a measure of distance between $\mathbf{X}_{i.t}$ and $\mathbf{X}_{l.t}$, $t \in T$;

VELOCITIES Distance

$$ {}_2\delta(i,l) = \pi_2 \sum_{t=1}^{T-1} \left\| \mathbf{V}_{i.t,t+1} - \mathbf{V}_{l.t,t+1} \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{V}_{..t,t+1}}} = \pi_2 \sum_{t=1}^{T-1} tr\left[\left(\mathbf{V}_{i.t,t+1} - \mathbf{V}_{l.t,t+1} \right)' \boldsymbol{\Sigma}^{-1}_{\mathbf{V}_{..t,t+1}} \left(\mathbf{V}_{i.t,t+1} - \mathbf{V}_{l.t,t+1} \right) \right] \quad (4) $$

where π_2 is a suitable weight to normalize the velocity dissimilarity, and $\boldsymbol{\Sigma}_{\mathbf{V}_{..t,t+1}}$ is the dispersion matrix of $\mathbf{V}_{..t}$.

Matrix $\boldsymbol{\Sigma}^{-1}_{\mathbf{V}_{..t,t+1}}$ allows to measure the autocorrelation between time points t, t+1

differences between velocities of objects i and l, in a time interval, are evaluated according

a measure of distance between $\mathbf{v}_{i.t,t+1} = (v_{i1t,t+1},\ldots,v_{ikt,t+1})'$ and $\mathbf{v}_{l.t,t+1}$, $t = 1,\ldots, u - 1$;

ACCELERATIONS distance

$$ {}_3\delta(i,l) = \pi_3 \sum_{t=1}^{T-2} \left\| \mathbf{A}_{i.t,t+2} - \mathbf{A}_{l.t,t+2} \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{A}_{..t,t+2}}} = \pi_3 \sum_{t=1}^{T-2} tr\left[\left(\mathbf{A}_{i.t,t+2} - \mathbf{A}_{l.t,t+2} \right)' \boldsymbol{\Sigma}^{-1}_{\mathbf{A}_{..t,t+2}} \left(\mathbf{A}_{i.t,t+2} - \mathbf{A}_{l.t,t+2} \right) \right] \, (5) $$

where π_3 is a suitable weight to normalize the acceleration dissimilarity, and $\boldsymbol{\Sigma}^{-1}_{\mathbf{A}_{..t,t+2}}$ is the dispersion matrix of $\mathbf{A}_{..t}$ that allows to measure the autocorrelation between time points t , t+2.

differences between accelerations of objects i and l, in a time interval, are evaluated according to a measure of distance between $\mathbf{a}_{i.t,t+2} = (a_{i1t,t+2},\ldots,a_{ikt,t+2})'$ and $\mathbf{a}_{l.t,t+2}$, $t = 1,\ldots, u - 2$.

3. Dissimilarities between trajectories

A dissimilarity between ***trends***, ***velocities*** *and* ***accelerations*** *of Y(i) and Y(l) is a*

mapping respectively from:

trends distances, i.e.: $\{{}_1\delta_t(i,l), t \in T\}$ *to* $\Re^+$*; velocities distances, i.e.:* $\{{}_2\delta_{t,t+1}(i,l), t \in T\}$ *to* $\Re^+$ *; accelerations distance, i.e.:* $\{{}_3\delta_{t,t+2}(i,l), t \in T\}$ *to* $\Re^+$.

The additive mapping has been considered:

$$d(i,l) = \pi_1 \sum_{t=1}^{T} \left\| \mathbf{X}_{i.t} - \mathbf{X}_{l.t} \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{X}_{..t}}} + \pi_2 \sum_{t=1}^{T-1} \left\| \mathbf{V}_{i.t,t+1} - \mathbf{V}_{l.t,t+1} \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{V}_{t,t+1}}} \quad \pi_3 \sum_{t=1}^{T-2} \left\| \mathbf{A}_{i.t,t+2} - \mathbf{A}_{l.t,t+2} \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{A}_{t,t+2}}} \quad (6)$$

4. Optimization problem

In a paper in progress Gorfarb, Summa and Vichi are clustering trajectories in a reduced space by using the T3CLUS Model (Rocci and Vichi, 2005).

The T3CLUS is a clustering version of the well known Tucker3 (T3) model proposed by Tucker (1966)

$$\mathbf{X}_{n,KT} = \mathbf{U}\,\overline{\mathbf{X}}_{G,KT}\ (\mathbf{CC}'\otimes\mathbf{BB}') + \mathbf{E}_{n,KT}\,. \quad (7)$$

Since occasions refer to time, we do not suppose to synthesize it by means of components;

thus, this dimension will remain unreduced.

This implies that in the previous model matrix **C** is an identity matrix of order *T*, i.e.,

$$\mathbf{X}_{n,KT} = \mathbf{U}\,\overline{\mathbf{X}}_{G,KT}\ (\mathbf{I}\otimes\mathbf{BB}') + \mathbf{E}_{n,KT}\,. \quad (8)$$

This model can be rewritten in frontal slabs form

$$\mathbf{X}_{..t} = \mathbf{U}\,\overline{\mathbf{X}}_{..t}\,\mathbf{BB}' + \mathbf{E}_{..t}. \quad (9)$$

The estimation of the model according to the distance between trajectories is a

Non-Ordinary LS problem

$$\pi_1 \sum_{t=1}^{T} \left\| \mathbf{X}_{..t} - \mathbf{U}\overline{\mathbf{X}}_{..t}\mathbf{BB}' \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{X}_{..t}}} + \pi_2 \sum_{t=1}^{T-1} \left\| \mathbf{V}_{..t,t+1} - \mathbf{U}\overline{\mathbf{V}}_{..t,t+1}\mathbf{BB}' \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{V}_{t,t+1}}} + \pi_3 \sum_{t=1}^{T-2} \left\| \mathbf{A}_{..t,t+2} - \mathbf{U}\overline{\mathbf{A}}_{..t,t+2}\mathbf{BB}' \right\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathbf{A}_{t,t+2}}} \rightarrow \quad \text{Min}$$

$$\mathbf{U},\mathbf{B}\ \overline{\mathbf{X}}_{..t},\ \overline{\mathbf{V}}_{..t,t+1},\ \overline{\mathbf{A}}_{..t,t+2}$$

subject to

[P1]

$\mathbf{B}'\mathbf{B}=\mathbf{I}_K$

U binary and row stochastic.

5. Partitioning Models

In the case the data are dissimilarities **D** the additive mapping (6) can be used, thus, in this case objects need to be partitioned from dissimilarity data. Three models can be used (Vichi, 2009).

WELL STRUCTURED PERFECT MODEL (Figure1)

$$\mathbf{D} = \mathbf{P}(\alpha_2(\mathbf{1}_n\mathbf{1}_n' - \mathbf{MM}') + \alpha_1(\mathbf{MM}' - \mathbf{I}_n))\mathbf{P}' + \mathbf{E} \qquad (10)$$

where:
D observed dissimilarity matrix
P WSP **classification matrix**
$\mathbf{1}_n$ is a vector of n ones,
$\mathbf{I}_n$ is the identity matrix
M is a ($n \times K$) matrix,
binary and row-stochastic,
α_1 ***heterogeneity within classes***
α_2 ***isolation between classes***
$0<\alpha_1\leq\alpha_2$.
E error matrix

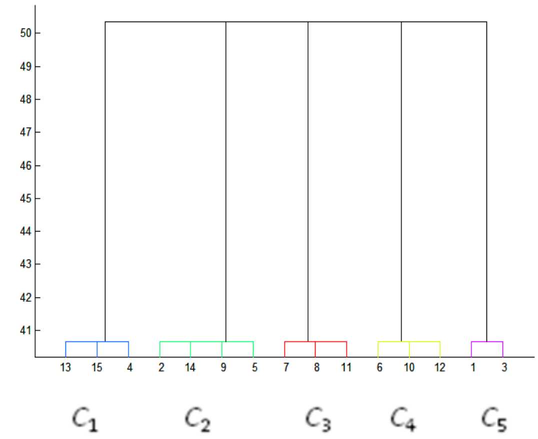


Figure1: Well structured perfect clustering model

WELL STRUCTURED CLASSIFICATION MODEL AND SQUARE K-MEANS (Figure 2)

$$\mathbf{D} = \overbrace{\mathbf{MD}_\mathrm{B}\mathbf{M}' + \mathbf{MD}_\mathrm{W}\mathbf{M}' - diag(\mathbf{MD}_\mathrm{W}\mathbf{M}')}^{\mathbf{Q}} + \mathbf{E} \qquad (11)$$

where:
D observed dissimilarity matrix
Q WS **classification matrix**
$\mathbf{1}_n$ vector of n ones,
$\mathbf{I}_n$ the identity matrix of order n and
M is a ($n \times K$), binary and row-stochastic,
$\mathbf{D}_\mathrm{W}$ matrix of heterogeneity within classes
$\mathbf{D}_\mathrm{B}$ matrix of isolation between classes
E error matrix
Square K-means matrix
R is not a classification matrix

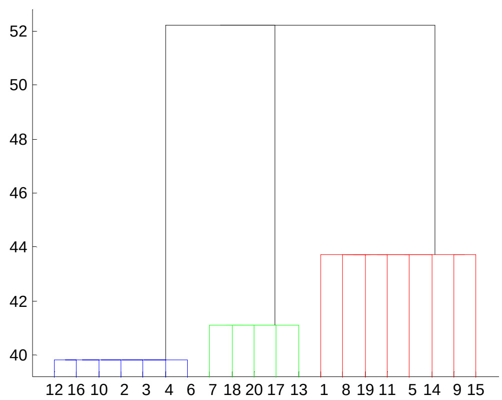


Figure2: Well structured classification model

$$\mathbf{D} = \overbrace{\mathbf{MD_BM}' + \mathbf{MD_WM}' - diag(\mathbf{MD_WM}')}^{\mathbf{Q}} + \mathbf{E} \qquad (12)$$

Subject to

$\mathbf{D_B}$ is *ultrametric matrix*

$max\ \mathbf{D_W} \leq \min \mathbf{D_B}$

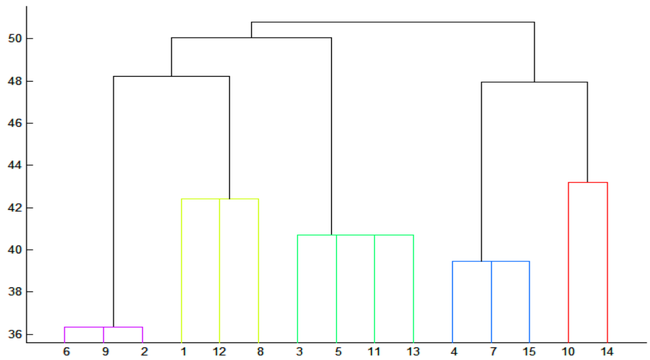


Figure 3: Example of Hierarchical partitioning model

References

ROCCI R, VICHI M. (2005). Three-mode Component Analysis with Crisp or Fuzzy Partition of Units. PSYCHOMETRIKA. vol. 70 n 4, pp. 715-736 ISSN: 0033-3123.

TUCKER L.R. (1966). Some mathematical notes on three-mode factor analysis. PSYCHOMETRIKA, *vol* 31, 279-311.

VICHI M., ROCCI R, H.A.L.KIERS. (2007). Simultaneous Component and Clustering models for three-way data: Within and Between Approaches. JOURNAL OF CLASSIFICATION. vol. 24 n1, pp. 71-98 ISSN: 0176-4268.

VICHI M. (2008). Fitting Semiparametric Clustering Models to Dissimilarity Data, ADVANCES IN DATA ANALYSIS AND CLASSIFICATION, vol 2, 2, 121-161,ISSN:1862-5347 .