Clustering models for high dimensional, temporal, and dissimilarity data

Maurizio Vichi  
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 \( X_1, X_2, \ldots, X_K \) 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

\[
\mathbf{Y} = \{ \mathbf{y}_{ijt} = (x_{ijt}, x_{i,jt}, \ldots, x_{i,kt}, t) : i \in I, t = 1, \ldots, T \}
\]

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

For each object \( i \), \( Y(i) = \{ y_{ijt} : t = 1, \ldots, T \} \) 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 \( y_{ijt} \) 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.:

$$v_{ij,t+1} = \frac{x_{ij,t+1} - x_{ij,t}}{s_{t,t+1}}. \quad (1)$$

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

In particular: $v_{ij,t+1} > 0$ ($v_{ij,t+1} < 0$) 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 $t+1$; $v_{ij,t+1} = 0$ 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 $a_{ij,t+2}$, is, for the $j$-th variable:

$$a_{ij,t+2} = \frac{v_{ij,t+1,t+2} - v_{ij,t+1,t+1}}{s_{t,t+2}}. \quad (2)$$

In particular: $a_{ij,t+2} > 0$ ($a_{ij,t+2} < 0$) if the object $i$, for the $j$-th variable, presents an increasing (decreasing) variation of velocity in the time interval from $t$ to $t+2$; $a_{ij,t+2} = 0$ 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

\[ d(i,l) = \pi_1 \sum_{t=1}^{T} \left[ \left( X_{i,t} - X_{l,t} \right) \right] = \pi_1 \sum_{t=1}^{T} tr\left[ \left( X_{i,t} - X_{l,t} \right) \Sigma_{X_{i,t}}^{-1} \left( X_{i,t} - X_{l,t} \right) \right] \]  

(3)

where \( \pi_1 \) is a suitable weight to normalize distances and \( \Sigma_{X_{i,t}} \) is the dispersion matrix of \( X_{i,t} \). Differences between trend intensities, in a time point \( t \), of objects \( i \) and \( l \) are evaluated according to a measure of distance between \( X_{i,t} \) and \( X_{l,t} \), \( t \in T \).

VELOCITIES Distance

\[ d_v(i,l) = \pi_2 \sum_{t=1}^{T} \left[ \left( V_{i,t,t+1} - V_{l,t,t+1} \right) \right] = \pi_2 \sum_{t=1}^{T} tr\left[ \left( V_{i,t,t+1} - V_{l,t,t+1} \right) \Sigma_{V_{i,t,t+1}}^{-1} \left( V_{i,t,t+1} - V_{l,t,t+1} \right) \right] \] 

(4)

where \( \pi_2 \) is a suitable weight to normalize the velocity dissimilarity, and \( \Sigma_{V_{i,t,t+1}} \) is the dispersion matrix of \( V_{i,t} \).

Matrix \( \Sigma_{V_{i,t,t+1}}^{-1} \) 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 to a measure of distance between \( V_{i,t,t+1} = (v_{i,t+2}, ..., v_{i,T})' \) and \( V_{l,t,t+1} = (v_{l,t+2}, ..., v_{l,T})' \), \( t = 1, ..., u - 1 \);

ACCELERATIONS distance

\[ d_a(i,l) = \pi_3 \sum_{t=1}^{T} \left[ \left( A_{i,t,t+2} - A_{l,t,t+2} \right) \right] = \pi_3 \sum_{t=1}^{T} tr\left[ \left( A_{i,t,t+2} - A_{l,t,t+2} \right) \Sigma_{A_{i,t,t+2}}^{-1} \left( A_{i,t,t+2} - A_{l,t,t+2} \right) \right] \]  

(5)

where \( \pi_3 \) is a suitable weight to normalize the acceleration dissimilarity, and \( \Sigma_{A_{i,t,t+2}}^{-1} \) is the dispersion matrix of \( A_{i,t} \) 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 \( a_{i,t,t+2} = (a_{i,t+2}, ..., a_{i,T})' \) and \( a_{l,t,t+2} = (a_{l,t+2}, ..., a_{l,T})' \), \( t = 1, ..., u - 2 \).

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.: \( \{ \delta_{i}(i, l), t \in T \} \) to \( \mathbb{R}^+ \); velocities distances, i.e.: \( \{ \delta_{i, t+1}(i, l), t \in T \} \) to \( \mathbb{R}^+ \); accelerations distance, i.e.: \( \{ \delta_{i, t+2}(i, l), t \in T \} \) to \( \mathbb{R}^+ \).

The additive mapping has been considered:

\[
d(i, l) = \pi_1 \sum_{t=1}^{T} \sqrt{X_{i,t} - X_{i,t+1}}^2 + \pi_2 \sum_{t=1}^{T-1} \sqrt{V_{i,t,t+1}}^2 + \pi_3 \sum_{t=1}^{T-2} \sqrt{A_{i,t,t+2}}^2
\]

(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)

\[
X_{n,KT} = U_{n,KT} (CC' \otimes BB') + E_{n,KT}.
\]

(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.,

\[
X_{n,KT} = U_{n,KT} (I \otimes BB') + E_{n,KT}.
\]

(8)

This model can be rewritten in frontal slabs form

\[
X_{i,t} = U_{i,t} BB' + E_{i,t}.
\]

(9)

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

Non-Ordinary LS problem

\[
\pi_1 \sum_{t=1}^{T} \|X_{i,t} - U_{i,t} BB\|_{w_{i,t}}^2 + \pi_2 \sum_{t=1}^{T-1} \|V_{i,t,t+1} - UVU_{i,t+1} BB\|_{w_{i,t,t+1}}^2 + \pi_3 \sum_{t=1}^{T-2} \|A_{i,t,t+2} - UVA_{i,t+2} BB\|_{w_{i,t,t+2}}^2 \rightarrow \text{Min}
\]

subject to

\[
U, B \text{ binary and row stochastic.}
\]
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 (Figure 1)**

\[
D = \alpha_2 \left( 2 \left( \text{I}_n \right)^T \text{I}_n \right)^{\frac{1}{2}} \left( I - MM^T \left( I_p \right) \right) + \left( 2 \left( \text{I}_n \right)^T \text{I}_n \right)^{\frac{1}{2}} \left( I_p \right) + E
\]

(10)

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

![Figure 1: Well structured perfect clustering model](image1)

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

\[
D = \left( \text{MD}_W M \right) + \left( \text{MD}_W M \right) \left( I - \text{diag} \left( \text{MD}_W M \right) \right) + E
\]

(11)

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

![Figure 2: Well structured classification model](image2)
**HIERARCHICAL PARTITIONING**

\[
D = \frac{q}{MD_B M} + \frac{1}{MD_W M} - \text{diag}(MD_W M) + E
\]  

Subject to

\(D_B\) is ultrametric matrix

\[\max D_W \leq \min D_B\]

![Figure 3: Example of Hierarchical partitioning model](image)

**References**


