Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(4): 396-411

Systematic Review

Longitudinal Data Clustering Methods: A Systematic Review

Arefeh Dehghani Tafti[1], Yunes Jahani[1,2], Sara Jambarsang[3], Abbas Bahrampour[1,2,4*]

[1]Department of Biostatistics and Epidemiology, Faculty of Public Health, Kerman University of Medical Sciences, Kerman, Iran.
[2]Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.
[3]Center for Healthcare Data Modeling, Department of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran.
[4]Griffith University, Brisbane, QLD, Australia.

ARTICLE INFO

Received 24.03.2023
Revised 11.04.2023
Accepted 26.04.2023
Published 15.12.2023

Key words:
Clustering;
Longitudinal data;
Non-parametric methods;
Model-based methods.

ABSTRACT

Introduction: In the last few decades, in many research fields, different methods were introduced to discover groups with the same trends in longitudinal data. The clustering process is an unsupervised learning method, which classifies longitudinal data based on different criteria by performing algorithms. The current study was performed with the aim of reviewing various methods of longitudinal data clustering, including two general categories of non-parametric methods and model-based methods.

Methods: In this research, to obtain related scientific articles, PubMed, Science Direct Scopus, ISI, and Google Scholar were searched between 2000 and 2021.

Results: According to our systematic review, the non-parametric k-means Clustering Method utilizing Euclidean distance emerges as a leading approach for clustering longitudinal data.

Conclusion: This research, with an overview of the studies done in the field of clustering, can help researchers as a toolbox to choose various methods of longitudinal data clustering in idea generation and choosing the appropriate method in the classification and analysis of longitudinal data.

Introduction

In few recent decades, longitudinal studies had different applications in many fields like medicine and Para medicine. The main characteristic of longitudinal data (LD) is that observations are measured frequently over time, resulting in a sequence of individuals that are usually related to each other. Longitudinal studies assist researchers in assessing how desired variables change over time. As data collection and storage capabilities continue to improve, longitudinal studies are being designed to incorporate numerous repeated measurements of a single variable for each subject over an extended period of time. Insights from behavior over time separate LD from other types of data. However, LD

*.Corresponding Author: abahrampour@yahoo.com, abahrampour121@gmail.com, a_bahrampour@kmu.ac.ir.

requires unique modeling approaches due to the correlation between measurements in each individual over time.

Clustering serves as a method within the realm of data mining to analyze data, and is performed with two main purposes:[1] the data in each cluster should be as similar as possible, in other words, the similarity within the clusters should be high,[2] the data in each cluster should be different from other clusters, so that the similarity between the clusters is low.[1]

lately, there has been notable advancement in the development of statistical techniques for analyzing LD in the clustering domain,[2-6] Specifically, several criteria are presented to determine the clusters optimal number,[7-9] and in addition for the missing values problem with standard assignment methods.[10] Many applications are proposed in this regard, including social and behavioral sciences to biomedicine,[7, 11-14] as well as new methods.[15-17]

Today, Some of these approaches can be classified into two main types: model-based (MB) and non-parametric (NP) methods.[18]

NP clustering algorithms, also known as traditional approaches, differ from other methods in that they make no assumptions about how the data were generated. Instead, they solely concentrate on determining the similarity between clusters and individuals. Their primary emphasis is on the heterogeneity criterion, the clusters number ,and the clustering algorithm type.[18]

In MB methods, the raw data is modeled as a combination of probability distributions through a standard statistical approach. The MB method is called as mixed model clustering. MB clustering has an extensive background, and in 1955, Tiedman introduced the notion of clusters as a fundamental component in the mixture model (MM) framework.[19] in the study published by Wolff In 1965, the evolution of MB clustering was investigated using Gaussian mixture model (GMiM).[20]

In recent years, researchers have developed various model-based clustering approaches for count data. For instance, Subedi and Browne (2020) proposed a novel approach that utilizes multivariate Poisson-log normal component distributions.[21] Roick et al. (2021) introduced a clustering method based on integer-valued autoregressive (INAR) models, which can effectively handle count data.[22] Ng and Murphy (2021) developed a model-based clustering approach for count process data, leveraging the Gaussian Cox process.[23] Additionally, Murphy et al. (2021) presented a distance-based mixture model for clustering life-course trajectory data.[24]

In the context of discrete data, including count-valued data, Karlis (2019) provided an overview of the key concepts and methods.[25] Furthermore, Salter-Townshend et al. (2012), and Bouveyron et al. (2019) reviewed various model-based clustering approaches for network data, highlighting their strengths and limitations.[26, 27] These studies demonstrate the importance of model-based clustering in understanding complex data structures.

In MB clustering algorithms, the values of the model's parameters are typically determined through Maximum Likelihood Estimation (MLE) and the Expectation Maximization (EM) algorithm.

In this article, we review and discuss various LD clustering methods. The methods were carefully chosen based on their common use and were specifically selected to offer a range of advantages and disadvantages. Our primary focus was on methods that can effectively

identify single-variable longitudinal patterns of change.

This research, with an overview of studies in the field of longitudinal data clustering methods, can help researchers as a toolbox in generating ideas and choosing the appropriate method in the classification and analysis of longitudinal data.

Methods

Paper Search Methods with Systematic Review

In this research, to obtain related scientific articles, PubMed, Science Direct Scopus, ISI, and Google Scholar were searched between 2000 and 2021 using the keywords longitudinal data, clustering, Unsupervised learning, non-parametric clustering, model-based clustering, single-variable, and their combinations.

Inclusion Criteria

The articles selected for review in this study included articles that specifically addressed LD clustering, research conducted between 2000 and 2021, the availability of the full text of the article, and the article having an appropriate structure.

After the investigations, the articles published at conferences and congresses and the articles that were published only on the websites, as well as the articles that were not of good quality in terms of content, were removed.

Exclusion Criteria

full texts of the included articles were studied and evaluated separately by two authors (RS and HFA) and in case of disagreement between their evaluation results, the third author (FM) announced the final opinion.

Among the reviewed studies, we excluded those studies that included at least one of the following criteria: 1) the type of clustering method was not mentioned, 2) the parameters of the proposed model were not clearly defined, 3) And also articles that were not of good quality in terms of content.

Study Selection

We identified 152 articles through a comprehensive database search. After removing duplicate articles, 61 articles remained. Then, irrelevant articles were removed after reviewing the titles, abstracts and full texts. After considering the exclusion criteria, 15 articles were included in the current systematic review. The study selection process is shown in Figure 1.

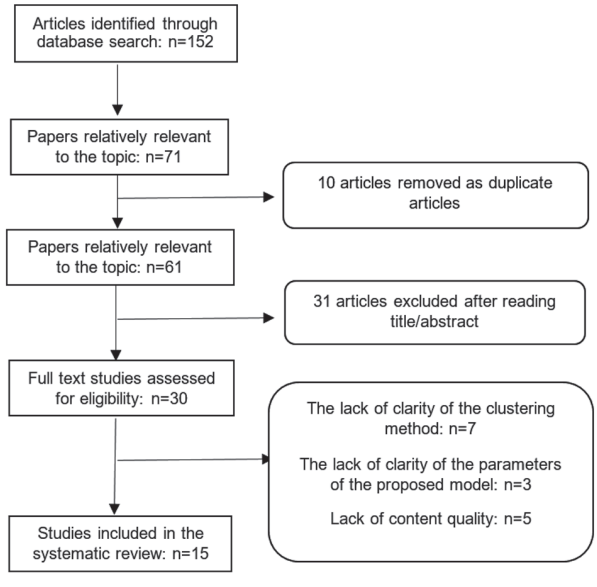

Figure 1. Study selection flowchart

Result

In this research, longitudinal clustering methods are grouped into two approaches: NP and MB methods

Non-parametric Clustering Methods (NPCM)

k-means Clustering Method for LD (KmL)

k-means Clustering Method for LD(KmL) is an approach for cluster analysis with the purpose to division n observations in k clusters (k ≤ n) whither every observation is allocated to the cluster whose centroid is closest.[28]

KmL is a hill-climbing algorithm that is introduced as a special case of the EM algorithm for iterative convergence.[29, 30] To utilize this technique, you generally need to specify the clusters number (k) as an input prior to initialization. To achieve the best division, the algorithm alternates among two stages: maximization (M) and waiting (E). These two steps are repeated until stabilization is achieved in cluster assignment. To formulate the KmL algorithm, it is assumed that there is a data set with n observations $X = \{x_1, \ldots, x_n\}$.[31, 32] Clustered in *k* groups $C = \{c_k . K = 1. \cdots . k\}$. In LD, every observation (X_i) shows a trajectory which is created by values of i^{th} observer in different times *{l = 1. … .t} and is shown by* $X_i = \{x_{i1} . x_{i2} . \cdots . x_{it}\}$. x_{it} is the measured value of i^{th} subject in the time *t*.

Essentially, the objective of KmL is to minimize the sum distance between every observation and central point of its corresponding cluster. Suppose that Z_k, which is the mean of the observations belonging to the corresponding cluster, represents the center of the cluster C_k. The square of the distance between Z_k and all observations X_i in the cluster Ck can be determined as the follows:

$$SD(C_K) = \sum_{x_i \in C_k} x_i - z_k^{\,2} \quad (1)$$

The KmL algorithm aim is to minimize the sum of squared distances among every observation and corresponding center in all k clusters.[33]

$$\sum_{k=1}^{K} \sum_{x_i \in C_k} x_i - z_k^{\,2} \quad (2)$$

k-means Clustering Method for LD Using Euclidean Distance

The commonly preferred metric for continuous variables is Euclidean distance (ED). Applying KmL with this type of distance can also be introduced as the traditional or usual approach.[31, 32]

The ED among the two paths x_1 and x_2 is obtained by the following equation:

$$d(x_1 . x_2) = x_1 - x_2 = \sqrt{\sum_{l=1}^{t} (x_{1l} - x_{2l})^2} \quad (3)$$

k-means Clustering Method for LD using the Manhattan Distance(MaD)

The Manhattan distance(MaD) is known as the "city block distance" or L1 distance. MaD is the absolute difference between points.[37]

The formula for calculating the MaD between two paths x_1 and x_2 is as follows:

$$d(x_1 . x_2) = \sum_{l=1}^{t} |x_{1l} - x_{2l}| \quad (4)$$

k-means Clustering Method for LD Using Chebychev Distance(CD)

The Chebychev distance (CD) between two vectors is their largest difference in any

coordinate dimension. It is named after Pafnuty Chebychev. CD is known as the maximum absolute value difference.[38]
The formula for calculating the CD between two paths x_1 and x_2 is as follows:

$$d\left(x_1.x_2\right) = max_l \left|x_{1l} - x_{2l}\right| \quad (5)$$

k-means Clustering Method for LD using Mahalanobis Distance(MD)

This approach assumes that we have a multivariate longitudinal response that takes values in p-dimensional space (R^p). With more precision, xil denotes a vector of length p containing p values of the response variables of continuous type, every observed at time *l=1. … .t.*
MD between the two paths x_1 and x_2 of the member $R^{(p\times t)}$ is obtained by the following equation:[39]

$$d_M\,(X_1.X_2)=(X_1-X_2)^T\,\Sigma^{-1}(X_1-X_2) \quad (6)$$

Where $\Sigma\epsilon R^{(p\times t)\times(p\times t)}$, and a diagonal block matrix is defined as follows:

$$\Sigma=\begin{bmatrix}\Sigma_1 & 0 & \cdots & 0\\ 0 & \Sigma_2 & & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \Sigma_p\end{bmatrix} \quad (7)$$

Assuming that the variance-covariance matrix among various time samples for variable k is as Σ_k *(k=1.2…..p).*

k-means Clustering Method for LD using Minkowski Distance(MinD)

Minkowski distance (MinD) is defined as a metric generalization of Manhattan and Euclidean distance. The formula for calculating the MinD among two trajectories x_1 and x_2 is defined as follows[40]:

$$d\left(x_1.x_2\right) = \left(\sum_{l=1}^{t}\left|x_{1l} - x_{2l}\right|^{\frac{1}{p}}\right)^p \quad (8)$$

In the above relationship, when p=2, the distance becomes ED, when p=1, the distance becomes MaD, and when p=∞, the distance becomes CD.

k-means Clustering Method for LD using Fréchet Distance (according to shapes of the trajectories)

Fréchet distance (FD) was introduced by Morris in 1906.[34] This similarity measure is used for geometric shapes and unlike traditional Euclidean distance, this method treats each trajectory as a curved path and determines the clusters based on the shape of the paths instead of their classical distance. For the first time, the algorithm of this distance type was presented by Alt and Godau in 1995.[35]
Formally, according to two definitions: (1) a reparameterization α of *[0,1]*, a continuous function without decrease spanning *α:[0,1] → [0,1]* with the condition *α(0) = 0* and *α(1) = 1.* (2) Consider a metric space S where a curve f in S is a continuous mapping from the unit interval *[0,1]* to S.
Consider two given curves f and g located in S. FD between two curves f and g in mathematical writing is defined as follows[10, 35, 36]:

$$\delta_F(f.g) = \text{infmax}\ \{f\,(á\,(t)) - g(á\,(t))\} \quad (9)$$
$$\alpha.\beta\ t\epsilon[0.1]$$

where ||.|| is the corresponding norm and is usually the Euclidean norm, and α and β are re-parameterized *[0,1]*.

Hierarchical Clustering Method (HCM)

The Hierarchical clustering method (HCM) is used to perform cluster analysis which, does not involve determining clusters number in the initial step. This algorithm usually clusters data based on distances.

Certain Hierarchical clustering (HC) algorithms utilization alternative clustering techniques, such as graph or density, as a auxiliary tool to build hierarchies.[36] In this method, two ED and dynamic time warping (DTW) are usually used.[37]

In the agglomerative HCM first, each observation is considered as a separate cluster, then at each step, the clusters that are more similar to each other are merged and create a larger cluster until finally all the observations are placed in one cluster. The reverse of this process occurs in a divisive HCM, so that first all observations are considered as a cluster, and in the next step, this cluster is divided into smaller clusters, and this process continues until only one observation is placed in each cluster.[38]

Model-based Clustering Methods (MB)

In this method, a statistical distribution is considered for the data. In These models, known as limited MMs, the data is assumed to be generated by a combination of probability distributions That each component display a distinct cluster. Therefore, when the data fits the model, it should be expected to perform well.

A finite mixed model is defined as follows:[39, 40]

$$f\left(x|\vartheta\right)=\sum_{k=1}^{K}\pi_k f_k\left(x\,|\,\theta_k\right) \qquad (10)$$

Where $x_1,\ldots,x_n$ are the observations of the independent random sample, and $X_1,\ldots,X_n$ is defined by parametric vector $\Theta=(\Theta_1,\ldots,\Theta_n)$. $f(xi;\theta)$ is called the mixed density function of k components. The component refers to the subgroups that make up the society, whose number is indicated by k, and Θk is the parameter related to the kth component, and αk is the *kth* weighting coefficient or the mixed coefficient ,and ɪk is the probability of an observation belonging to the *kth* component with the conditions $0\leq\pi k\leq 1$ and $\sum_{\pi k}=1$.

Gaussian Mixture with Cholesky Decomposition

The density function of the GMM is in the form of the following model:[39]

$$f(x|\vartheta)=\sum_{k=1}^{K}\pi_k\emptyset(x|\mu_k.\sum_k) \qquad (11)$$

Where, π_k represents the probability of membership in the *Kth* group. $\emptyset\,(x|\mu_k,\sum_k)$ is the density function of a multivariate Gaussian (MG) distribution including the mean of μk and matrix of $\sum_k$.

In this method, the Gaussian mixed model with the Cholesky decomposed covariance structure is used for every mixed component. According to the modified Cholesky decomposition, the covariance matrix can be decomposed into expressions $T\sum T'=D$, whither D presents a unique diagonal matrix involve positive diagonal entries, and T presents a unique lower triangular matrix with diagonal elements 1. other method to formulate the modified Cholesky analysis is in the form of $\sum^{-1}=T'D\text{-}1T$, where

T and *D* values are presented as innovation variances values and generalized autoregressive coefficients, respectively. Therefore, according to the Cholesky decomposed covariance matrix, the density function of an observation xi in group k is given by the following function.[41]

$$f\left(x_i|\mu_k.T_k.D_k\right)=\frac{1}{\sqrt{(2\pi)^p\left|D_k\right|}}$$
$$exp\left\{-\frac{1}{2}\left(x_i-\mu_k\right)^{'}T_k^{'}D_k^{-1}T_k\left(x_i-\mu_k\right)\right\} \quad (12)$$

where T_K is the lower triangular matrix with $p*p$ dimension and D_K is the $p*p$ diagonal matrix that follows the modified Cholesky covariance matrix decomposition.

Gaussian Mixed-effects Model by Smoothing Spline Estimation

In this approach, a Gaussian mixed-effects model (GMEM) with NP spline smoothing is used. Suppose that $X=\{x_1. \ldots .x_n\}$ is a set of n observations, for every observation, $X_i=\{x_{i1}.x_{i2}. \ldots .x_{it}\}$ is a single trajectory that is measured at different time points $\{l=1.2.\ldots.t\}$. *N* observations a re classified in k groups based on their trajectory trends. In this method, the individual trajectory X_i is modeled as a linear combination of a fixed effect at time *t* , $\xi_k(l)$ related to cluster *k*, a random effect *βi*, and an error term ϵit:

$$x_{il}=\xi_k\left(l\right)+\beta_i+\dot{o}_{il} \quad (13)$$

where $\epsilon it \sim N(0,\Theta)$ and $\beta i \sim N(0,\Theta K)$. The fixed effect *ξk* corresponds to the trend of overall or baseline trajectories associated with cluster *k*. Any systematic change from the general trend is represented by the random effect *βi* and ϵ_{il} represents the measurement error. Therefore, *Xi* follows a multivariate normal distribution $N(\xi_k,\sum_k)$ with variance $\sum_k$, which is defined as the following relationship:[42]

$$\sum_k=\theta_k I_t+\theta J_t=\begin{pmatrix}\theta_k+\theta & \theta & \ldots & \theta\\ \theta & \theta_k+\theta & \ldots & \theta\\ \ldots & \ldots & \ldots & \ldots\\ \theta & \theta & \ldots & \theta_k+\theta\end{pmatrix} \quad (14)$$

where I_t is a unit mat rix with dimension *t* and J_t is a square matrix of one with dimension *t*. This issue of clustering i s projected in a mixed model where every cluster can be explained by a Gaussian distribution with $N(\xi_k,\sum_k)$ parameters:

$$x_i \sim \sum_{k=1}^{K}\pi_k\mathcal{N}(\xi_k.\Sigma_k) \quad (15)$$

where *πk* represents the mixed coefficients in the mixed model.

In this model, a NP method using a smoothing spline is used instead of a parametric approach to select the base value of *ξk*. The model tries to minimize the relation (9) by fitting a cubic spline *ξk* to a collection of observations and finding the appropriate *ξk*:

$$\sum_{l=1}^{t}\left(x_{il}-\xi_k\left(l\right)\right)^2+\lambda_k\int\left(\xi_k^{"}\left(t\right)\right)^2dt \quad (16)$$

Where $\sum (xit-\xi_k(t))^2$ is the quantification of values deviation from curve ξ_k, and $\lambda_k\int\xi$ penalized the curve's un-smoothness.

If variable x_{it} has a normal distribution, the first polynomial is fitted with the negative probability of entrance into the system ,and the curve is transformed into the following penalized probability:[42]

$$-L\left(x_i\right)+\lambda_k\int\left(\xi_k^{"}\left(t\right)\right)^2dt \quad (17)$$

Were $L(x_i)$ is the data log probability.

Bayesian Hierarchical Clustering

Bayesian Hierarchical Clustering (BHC) Similar to the traditional cumulative HCM allocates each observation to a specific cluster and repeatedly merges pairs of clusters, But with the difference that it uses statistical hypothesis testing to determine which clusters should be merged.[43] Suppose $X=\{X_1. \dots .X_n\}$ represents the data set and $D_i \subset x$ represents the set of observations in the leaves under the tree T_i. The initialization process begins with the creation of n distinct trees $\{T_i : i=1,\dots,n\}$, each including one observation. In each step of the algorithm, there is a thorough evaluation to explore all potential mergers of two trees.[44] To consider each integration, two hypotheses are proposed, the first hypothesis (H_{1k}) is that all the data in D_k are identically and directly generated from the same probable model $p(x|\Theta)$ with unknown parameters Θ. It is also assumed that this probable model is a MG model with parameters $\Theta=(\mu,\sum)$. The alternative hypothesis (H_{2k}) indicates that the *Dk* data has two or more clusters.

To appraise the probability of the data under the hypothesis (H_1^k), we need to specify the prior value under the model parameter $p(\Theta|\beta)$ with super parameters β. The probability of data D_k under (H_1^k) is equal to:[45]

$$p(D_k|H_1^k)=\int p(D_k|\theta)p(\theta|\beta)d\theta=\int[\prod_{xi\epsilon Dk}p(x_i|\theta)]p(\theta|\beta)d\theta \quad (18)$$

This relation calculates the probability that all D_k data are generated from parameter values, assuming a model of the from $p(x|\theta)$.

Under the alternative hypothesis (H_2^k), the probability of observing the given data (D_k) is calculated as the product of the probabilities of each individual data point (D_i) in the subtrees (T_i) and (T_j). This can be expressed mathematically as $p(D_k|H_2^k) =p(D_i|T_i)\, p(D_j|T_j)$. By compounding the probability of the data under the hypotheses H_1^k and H_2^k, weighted by the prior probability that all the points in D_k belong to the same cluster $\pi_k=p(H_1^k)$, the marginal probability of the data in the T_k tree is obtained as the following relationship:[45]

$$p(D_k|T_k)=\pi_k p(D_k|H_1^k)+(1-\pi_k)p(D_i|T_i)p(D_j|T_j) \quad (19)$$

Then the posterior probability of the integrated hypothesis $r_k=p(H_1^k|p_k)$ is calculated using the Bayes rule as the following relationship:[45]

$$r_k=\frac{\pi_k p\left(D_k|H_1^k\right)}{p\left(D_k|T_k\right)}=\frac{\pi_k p\left(D_k|H_1^k\right)}{\pi_k p\left(D_k|H_1^k\right)+\left(1-\pi_k\right)p\left(D_i|T_i\right)p\left(D_j|T_j\right)} \quad (20)$$

If $r_k > 0.5$, it indicates a higher probability that the data points in the trees are originating from a single underlying function, as a result, they converge and the tree can be felled at the points where $r_k < 0.5$ and the branches form separate clusters.

Latent Profile Analysis

Latent Profile Analysis (LPA) is a statistical approach that aims to determine unobserved subgroups within a population (i.e., profiles).[46] This method is usually also known as Latent Class Analysis (LCA).[47,48] In some cases, An LPA application is specifically known as longitudinal LPA (LLPA). The observation expected value at time t hinges on cluster membership According to the cluster g membership, we have:

$$x_{i.j} = \mu_{g.j} + \varepsilon_{g.i.j} \quad . \quad i \in I_g$$
$$\varepsilon_{g.i.j} \sim N\left(0.\sigma_{g.j}\right) \tag{21}$$

where $\mu_{g,j}$ and $\sigma_{g,j}$ respectively represent the specific mean of the cluster and the specific standard deviation of the cluster at time t_j. The observations probability density function of the *ith* subject is calculated by marginalization on all clusters of G:

$$f\left(x_i\right) = \sum_{g=1}^{G} \pi_g \prod_{j=1}^{n} \varnothing(x_{i.j} \mid \mu_{g.j}.\sigma_{g.j}) \tag{22}$$

where *φ(.)* represents the probability density function with distribution of normal and π_g represents the cluster ratio with the condition $\sum_{g=1}^{G} \pi_g = 1$ and $\pi_g > 0$. To reduce the parameters number, usually the variance between measurements over time is assumed to be equal ($\sigma_{g,j}=\sigma_g$).[44]

This model is generally estimated through MLE using the EM algorithm.[49] Here, the data according to the unknown observation model $\theta=(\pi_1. \ldots .\pi_{G-1}.\mu_1. \ldots .\mu_G.\sigma_1. \ldots .\sigma_G)$ and the unknown cluster membership matrix *z*, where $z_{i,g}$ is the probability of observation *i* belonging to cluster *g* conditional on *θ*.

Two-step approach using Growth Curve modeling and K-means (GCKM)

A two-stage clustering approach is demonstrated by modeling pathways using clustering subject parameter estimates and a growth curve model (GCM) (i.e., random effects) using the K-means method. The GCM is estimated in a MM framework. The GCKM model is designed to analyze LD sets by representing them using a single group trajectory (the fixed effects), and for each subject, their unique deviation from this trajectory (the random effects). The paths that an object follows, known as trajectories, are commonly represented by either a first-order or a second-order polynomial.[28,31,32]

The trajectory described in terms of polynomials of order *k* and random effects in all terms is determined by the following equation:

$$x_{i.j} = \sum_{k=0}^{K} \beta_{k.i} t_{i.j}^{k} + \varepsilon_{i.j} \tag{23}$$

$$\beta_{k.i} = \alpha_k + \zeta_{k.i} \tag{24}$$

Here, α_k denotes the coefficient of the *kth* order of the polynomial path, the notation εi,j represents the measurement error (intra-subject variability) and the notation $f_{k,i}$ represents the subject random effect i of for the *kth* coefficient (i.e. inter-subject variability). The random effects are assumed to have a multivariate distribution with zero mean and an unstructured variance-covariance matrix. These effects are independent of the measurement error(ε). It is also assumed that the measurement error is normally distributed and independently with zero mean and common variance.

The random effects $\varepsilon_{k,i}$ of each path are usually predicted using the best linear unbiased predictors (BLUPs) and they are input vectors $X_i = \left(\hat{\zeta}_{0i} .\hat{\zeta}_{1i} \ldots..\hat{\zeta}_{Ki}\right)$ transferred to the KmL algorithm.

Group-Based Trajectory Model (GBTM)

Group-Based Trajectory Model (GBTM) is usually introduced as latent-class growth analysis (LCGA, LCGM), sometimes as NP multilevel MMing (NPMM), and

semi-parametric group-based modeling (SGBM).[32,50,51]

The GBTM model describes a LD set based on a combination of group trajectories, regardless of within-group variability. Similar to the methods concept such as KmL or LPA, the GBTM model describes population heterogeneity through a set of homogeneous clusters in which subjects are represented only by the trajectory of their respective cluster.

The GBTM model describes trajectories using a linear model. For a given trajectory X_i, its observations are described by the trajectory group g as follows:

$$x_{i.j}^{(g)} = \sum_{k=0}^{K} \alpha_k^{(g)} t_{i.j}^k + \varepsilon_{i.j} \quad (25)$$

where ε_{ij} represents the remainder at time t_{ij} and $\alpha_k^{(g)}$ represents the kth coefficient for the polynomial of group g. In this context, the subject trajectories $\Psi_i(t_{ij})$ are all the same as $\sum \alpha_k^{(g)} t_{ij}^k$ when the subject *i* belongs to group g. It is assumed that the residuals $\varepsilon_{i.j}$ are independently distributed with a normal distribution with zero mean and variance σ_y^2. The marginal average of GBTM is calculated according to the following equation:[52]

$$\mathbb{E}\left(x_{i.j}\right) = \sum_{g=1}^{G} \pi^{(g)} \sum_{k=0}^{K} \alpha_k^{(g)} t_{i.j}^k \quad (26)$$

Growth Mixture (GMM)

Growth Mixture (GMM) is a generalization of the GBTM model by including parametric random effects and enables a better fit with the data assuming intra-cluster variability.[53, 54]

The GMM method is a generalization of the GBTM method by considering the coefficients $\alpha_k^{(g)}$ in equation (26) for a particular subject, which basically introduces a mixed- effects model in each group *g*. Therefore, a certain trajectory X_i is introduced by the group g in the form of the following relation:

$$x_{i.j}^{(g)} = \sum_{k=0}^{K} \beta_{k.i}^{(g)} t_{i.j}^k + \varepsilon_{i.j}^{(g)} \quad (27)$$

$$\beta_{k.i}^{(g)} = \alpha_k^{(g)} + \zeta_{k.i}^{(g)} \quad (28)$$

Group-dependent fixed effects are defined by $\alpha_k^{(g)}$.

In this model, it is assumed that the residuals follow an independent normal distribution with zero mean and uncorrelated with $\zeta_{kj}^{(g)}$.we assume that the random effects have a normal distribution with zero mean but it may be correlated in group *g* of course, independent of random effects between groups. The marginal mean of the GMM model is defined as the following relationship:

$$\mathbb{E}\left(x_{i.j}\right) = \sum_{g=1}^{G} \pi^{(g)} \sum_{k=0}^{K} \left(\alpha_k^{(g)} + \zeta_{k.i}^{(g)}\right) t_{i.j}^k \quad (29)$$

Where $0 < \pi^{(g)} \leq 1 . \sum \pi^{(g)} = 1 . E\left(\varepsilon_{k.j}^{(g)}\right) = 0$

Discussion

To raise awareness among authors about the application of clustering methods in longitudinal data, this study was conducted in the form of a systematic review. Specifically, the research focused on developing clustering methods for analyzing univariate longitudinal data. The primary goal of this investigation was to introduce and evaluate two non-parametric and model-based approaches for clustering longitudinal data. Based on the results of this study, we highlighted the strengths and limitations of each clustering method, presented insights into their potential applications and areas for further improvement.

model-based methods assume that a mixture of underlying probability distributions generates the data and that it can be described using a standard statistical model.[55-57] In model-based clustering algorithms, the parameters of each distribution are usually estimated by maximizing the likelihood. Thus, a particular clustering method can be expected to work well when the data fits the model.

Unlike model-based approaches, non-parametric clustering methods have no assumptions on how the data was generated and explicitly focus on defining similarity between subjects and clusters. They mainly focus on the dissimilarity measure, the clustering algorithm and the number of clusters. Non-parametric clustering methods may be referred to as the traditional approaches. The K-means algorithm is by far the most used non-parametric method and has already been extended and adapted to longitudinal data.[10, 29, 58]

Da Costa et al. concluded that the non-parametric KmL method with Euclidean distance and the hierarchical method with Euclidean distance were the best clustering methods, respectively, and KmL methods with MaD and MB methods such as the GMM with Kulsky variance-covariance matrix decomposition were also Good results were obtained.[59]

In Den Teuling et al.'s study, considering all the conditions of number of repeated observations, sample size, number of groups in the simulations, and within-group variability, the GMM and GCKM methods significantly outperformed the non-parametric GBTM methods From the point of view of estimation of group routes and group assignment, they perform better in all scenarios.[16]

According to the Study findings of Den Teuling et al., the difference in the performance of GCKM and KmL shows the advantage of reducing the dimensions In the initial stage and describes the Features of the path more briefly.[16] The results of this study are contrary The findings of Twisk et al. showed that GCKM and KmL provided similar results, in other words, GCKM and KmL methods have almost similar solutions in all scenarios.[47]

The study conducted by Feldman and his team revealed that Longitudinal Latent Class Analysis (LLCA), which can be regarded as a straightforward clustering method similar to KmL, yields outcomes comparable to GBTM.[48] Dentoling et al.'s findings showed that KmL has a better performance than GBTM due to significant flexibility in describing paths, better scaling, and less calculation time. However, GBTM method is preferred in data analysis containing missing or misaligned observations.[16]

The GMM model performs well in studies with a small number of observations, while in the GCKM model, increasing the number of observations allows a more accurate estimation of the random effects of the model. Because of equivalent outcomes of GMM for better execution time scaling with model complexity, and a larger number of observations, GCKM is the preferred choice for ILD due to its computational rapidity. However, under greater within-group variability, only GMM and GCKM were able to do this satisfactorily.[16]

The result of the simulation in the study of Den Tuling et al. showed that GCKM and GMM perform better than GBTM and KmL methods in the dataset including heterogeneous subgroups.[16] The GBTM method is more sensitive to outlying observations than the GMM method and has a faster algorithm execution, and also tends to overestimate the

clusters number.[60]

Over the past two decades, significant advances have been made in the development of longitudinal data clustering methods. While this has led to a wide range of techniques being explored, it is not possible to cover every topic in this article.

Instead, we have focused on introducing the fundamental and most influential methods, considering their strengths and limitations and with a particular emphasis on univariate longitudinal data analysis.

Conclusion

MB clustering techniques tend to require a relatively small sample size in terms of both the trajectories number and the observations number in each trajectory. In MB methods, unlike NP methods, the representation of the cluster path is parametric. NP methods are theoretically less complicated and have fewer software limitations. Although the NP methods have a high speed in calculation and access to the implementation of the algorithm widely, but they are highly sensitive to the measurement noise.

Declarations

Conflict of interest

The Authors have declared no conflict of interest.

Acknowledgments

All individuals and colleagues who helped us with their scientific efforts to develop this paper are sincerely appreciated.

References

1. Kaur Mann NKA. Review Paper on Clustering Techniques. Global Journal of Computer Science and Technology. 2013;13:42-7.

2. Abraham C, Cornillon P, Matzner-Lober E, Molinari N. Unsupervised Curve Clustering using B-Splines. Scandinavian Journal of Statistics. 2003;30(3):581-95.

3. Fitzmaurice G, Laird N, Ware J. Applied longitudinal analysis. 2 ed: Wiley; 2011.

4. James GM, Sugar CA. Clustering for Sparsely Sampled Functional Data. Journal of the American Statistical Association. 2003;98(462):397-408.

5. Rossi F, Conan-Guez B, Golli A. Clustering Functional Data with the SOM. 2004.

6. Tarpey T, Kinateder KK. Clustering functional data. Journal of classification. 2003;20(1).

7. Caliński T, Ja H. A Dendrite Method for Cluster Analysis. Communications in Statistics - Theory and Methods. 1974;3:1-27.

8. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 1985;50(2):159-79.

9. Yosung S, Jiwon C, In-Chan C, editors. A Comparison Study of Cluster Validity Indices Using a Nonhierarchical Clustering Algorithm.

International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06); 2005 28-30 Nov. 2005.

10. Genolini C, Ecochard R, Benghezal M, Driss T, Andrieu S, Subtil F. kmlShape: An Efficient Method to Cluster Longitudinal Data (Time-Series) According to Their Shapes. PloS one. 2016;11(6):e0150738.

11. Delmelle EC. Mapping the DNA of Urban Neighborhoods: Clustering Longitudinal Sequences of Neighborhood Socioeconomic Change. Annals of the American Association of Geographers. 2016;106(1):36-56.

12. Hedeker D, Gibbons RD. Longitudinal data analysis: Wiley-Interscience; 2006.

13. Morris R, Blashfield R, Satz P. Developmental classification of reading-disabled children. Journal of clinical and experimental neuropsychology. 1986;8(4):371-92.

14. Qin S, Jiao K, He J, Lyu D. Forage crops alter soil bacterial and fungal communities in an apple orchard. Acta Agriculturae Scandinavica, Section B — Soil & Plant Science. 2016;66(3):229-36.

15. Ciampi A, Campbell H, Dyachenko A, Rich B, McCusker J, Cole MG. Model-Based Clustering of Longitudinal Data: Application to Modeling Disease Course and Gene Expression Trajectories. Communications in Statistics - Simulation and Computation. 2012;41(7):992-1005.

16. Den Teuling NGP, Pauws SC, van den Heuvel ER. A comparison of methods for clustering longitudinal data with slowly changing trends. Communications in Statistics - Simulation and Computation. 2021:1-28.

17. Maruotti A, Vichi M. Time-varying clustering of multivariate longitudinal observations. Communications in Statistics-Theory and Methods. 2016;45(2):430-43.

18. Heggeseth BC. Longitudinal cluster analysis with applications to growth trajectories: University of California, Berkeley; 2013.

19. Tiedeman D, editor On the study of types. Symposium on pattern analysis: Air University, USAF School of Aviation Medicine Randolph Field, TX; 1955.

20. Wolfe J. A Computer Program for the Maximum-Likelihood Analysis of Types. 1965:57-60.

21. Subedi S, Browne RP. A family of parsimonious mixtures of multivariate Poisson-lognormal distributions for clustering multivariate count data. Stat. 2020;9(1):e310.

22. Roick T, Karlis D, McNicholas PD. Clustering discrete-valued time series. Advances in Data Analysis and Classification. 2021;15:209-29.

23. Ng TLJ, Murphy TB. Model-based clustering of count processes. Journal of Classification. 2021;38:188-211.

24. Murphy K, Murphy TB, Piccarreta

R, Gormley IC. Clustering longitudinal life-course sequences using mixtures of exponential-distance models. Journal of the Royal Statistical Society Series A: Statistics in Society. 2021;184(4):1414-51.

25. Karlis D. Mixture modelling of discrete data. Handbook of Mixture Analysis. 2019:193-218.

26. Bouveyron C, Celeux G, Murphy TB, Raftery AE. Model-based clustering and classification for data science: with applications in R: Cambridge University Press; 2019.

27. Salter-Townshend M, White A, Gollini I, Murphy TB. Review of statistical network analysis: models, algorithms, and software. 2012.

28. MacQueen J, editor Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability; 1967.

29. Genolini C, Falissard B. KmL: a package to cluster longitudinal data. Computer methods and programs in biomedicine. 2011;104(3):e112-21.

30. Celeux G, Govaert G. A classification EM algorithm for clustering and two stochastic versions. Computational Statistics & Data Analysis. 1992;14(3):315-32.

31. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. Biometrics. 1982;38(4):963-74.

32. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annual review of clinical psychology. 2010;6:109-38.

33. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010;31(8):651-66.

34. Fréchet MM. Sur quelques points du calcul fonctionnel. Rendiconti del Circolo Matematico di Palermo (1884-1940). 1906;22(1):1-72.

35. Alt H, Godau M. Computing the Fréchet Distance between Two Polygonal Curves. Int J Comput Geometry Appl. 1995;5:75-91.

36. Agarwal PK, Avraham RB, Kaplan H, Sharir M. Computing the Discrete Fréchet Distance in Subquadratic Time. SIAM Journal on Computing. 2014;43(2):429-49.

37. Bellman R, Kalaba R. On adaptive control processes. IRE Transactions on Automatic Control. 1959;4(2):1-9.

38. Han J, Kamber M, Pei J. Data Mining,Concepts and Techniques. 3rd Edition ed. Edition T, editor2012. 459-61 p.

39. McNicholas PD. Model-based clustering. Journal of Classification. 2016;33(3):331-73.

40. Martinez WL, Martinez AR. Computational statistics handbook with MATLAB. Edition r, editor: Chapman and Hall/CRC; 2015.

41. McNicholas PD, Murphy TB. Model-based clustering of longitudinal data. The

Canadian Journal of Statistics / La Revue Canadienne de Statistique. 2010;38(1):153-68.

42. Golumbeanu M, Beerenwinkel N, editors. Clustering time series gene expression data with TMixClust2018 20182018.

43. Heller KA, Ghahramani Z, editors. Bayesian hierarchical clustering. Proceedings of the 22nd international conference on Machine learning; 2005.

44. Peugh J, Fan X. Modeling Unobserved Heterogeneity Using Latent Profile Analysis: A Monte Carlo Simulation. Structural Equation Modeling: A Multidisciplinary Journal. 2013;20(4):616-39.

45. Savage RS, Heller K, Xu Y, Ghahramani Z, Truman WM, Grant M, et al. R/BHC: fast Bayesian hierarchical clustering for microarray data. BMC Bioinformatics. 2009;10(1):242.

46. Lazarsfeld PF, Henry NW. Latent Structure Analysis: Houghton, Mifflin; 1968.

47. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: A comparison between methods. Journal of clinical epidemiology. 2012;65:1078-87.

48. Feldman BJ, Masyn KE, Conger RD. New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. Developmental psychology. 2009;45(3):652-76.

49. McLachlan GJ, Peel D. Finite mixture models: John Wiley & Sons; 2004.

50. Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. Psychological Methods. 1999;4(2):139-57.

51. Nagin DS, Tremblay RE. Developmental trajectory groups: Fact or a useful statistical fiction? Criminology: An Interdisciplinary Journal. 2005;43(4):873-904.

52. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annual review of clinical psychology. 2010;6:109-38.

53. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics. 1999;55(2):463-9.

54. Verbeke G, Lesaffre E. A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population. Journal of the American Statistical Association. 1996;91(433):217-21.

55. Gong H, Xun X, Zhou Y. Profile clustering in clinical trials with longitudinal and functional data methods. Journal of biopharmaceutical statistics. 2019;29(3):541-57.

56. Schramm C, Vial C, Bachoud-Lévi A-C, Katsahian S. Clustering of longitudinal data by using an extended baseline: A new method for treatment efficacy clustering in longitudinal data. Statistical methods in medical research. 2018;27(1):97-113.

57. Zhu X, Qu A. Cluster analysis of

longitudinal profiles with subgroups. 2018.

58. Genolini C, Alacoque x, Sentenac M, Arnaud C. kml and kml3d : R Packages to Cluster Longitudinal Data. Journal of Statistical Software. 2015;65:1-34.

59. Da Costa JP, Garcia A. New confinement index and new perspective for comparing countries - COVID-19. Computer methods and programs in biomedicine. 2021;210:106346.

60. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. J Clin Epidemiol. 2012;65(10):1078-87.