Scope
Volume 14 Number 1 March 2024

A Comprehensive Review on Triclustering Techniques in Three-Dimensional Data Analysis: Unveiling Patterns Across Biomedical and Social Domains

Dr. Swathypriyadharsini P[a*], Dr. K. Premalatha[b]

[*a]Assistant Professor III, CSE department, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamil Nadu, India, [b]Professor, CSE department, Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamil Nadu, India,

Corresponding author: **Dr. Swathypriyadharsini P**

Abstract

Three-dimensional data are increasingly prevalent across biomedical and social domains. Notable examples are gene-sample-time, individual-feature-time, or node-node-time data, generally called observation attribute-context data. The unsupervised analysis of three-dimensional data can be pursued to discover putative biological modules, disease progression profiles, and communities of individuals with coherent behaviour, among other patterns of interest. It is thus key to enhancing the understanding of complex biological, individual, and societal systems. The clustering technique is one of the important unsupervised approaches for mining similar patterns either row-wise or column-wise. Biclustering performs simultaneous clustering of both rows and columns by identifying the similarities under a specific subset of conditions. On the other hand, the Triclustering algorithm extracts similar pattern subsets including row, column and also the third dimension mostly as time. This review paper focuses on the triclustering approach followed in many kinds of data such as binary data, big data and most importantly in gene expression data. This work also divulges the computational overhead in dealing the three-dimensional data. It also provides a detailed view of the approaches followed in different triclustering algorithms, measures used, dataset applied and also the validation framework followed. Finally, it highlights challenges and opportunities to advance the field of triclustering and its applicability to complex three-dimensional data analysis.

Keywords: Triclustering; Three-dimensional data; Gene Expression Data; Comparative Review; Clustering; Sub-space Clsutering;

1. Introduction

Bioinformatics is an interdisciplinary field that combines biological sciences with computational methods, playing a pivotal role for researchers and scientists engaged in various biological experiments aimed at enhancing the well-being of living organisms. It depends on three principal elements such as huge database with a vast array of biological data, algorithms, statistical techniques employed to elucidate relationships within expansive datasets, and computational tools designed for analysing and interpreting biological data. Within biological databases, Microarray gene expression data is one of the biological databases that contain the gene expression values for different samples. The primary objective of analyzing

microarray gene expression data is to identify patterns in gene expression that differentiate experimental and control samples, facilitating the classification of new samples based on their gene expression profiles.

1.1. Three-Dimensional Microarray Data

Microarray technology has been very effective in the examination of the expression of thousands of genes at a time and it has revolutionized the study of gene expression data. The activity of measuring all genes for the number of biological replicates across each space or time point is referred to as three-dimensional datasets. The time series datasets in microarray technology have been used to measure in a single experiment, the expression values of thousands of genes under a huge variety of experimental conditions across different time points. Due to its huge volume of data, several computational methods are needed to analyze such datasets. Fig. 1 shows the three-dimensional time series dataset representation in which each slice represents one-time point.

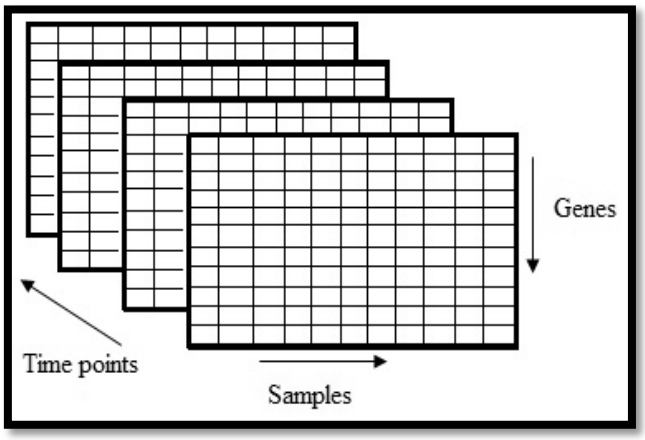

Fig. 1. Three-Dimensional Dataset Representation

1.2. Clustering Analysis

Data mining research mainly targets on effective and scalable knowledge discovery from databases that give appropriate solutions in time. Clustering is one of the unsupervised approaches for revealing the interesting patterns in the underlying data. Clustering algorithms group, the data objects into sets of disjoint classes known as clusters. It aims to maximize similarity within the clusters as well as to minimize similarity between the clusters, based on a distance measure (Jiang et al.2006). In the context of gene expression data, extracting information of gene expression levels that vary among the different conditions includes grouping of co-expressed genes. If two genes have similar expression profiles across some conditions or samples, then there may be some relationship between their functions which reflects a common regulation pattern (Pollard & Van Der Laan 2002). Traditional clustering algorithms may be either gene-based clustering or sample-based clustering. In gene-based clustering, the genes that are similar across a set of samples are clustered whereas in sample-based clustering, the samples that are similar across a set of genes are grouped. However, these clustering algorithms fail to find the group of genes that are similarly expressed over a subset of experimental conditions. This problem is solved by biclustering algorithms (Cheng & Church 2000).

1.3. Biclustering

Biclustering is a two-dimensional clustering method where the genes and conditions are grouped simultaneously. A bicluster can be defined as a subset matrix with a set of genes and a set of samples or experimental conditions. With a given gene expression matrix, a bicluster has a set of genes that behave similarly under a subset of experimental conditions. There are different types of biclusters extracted such as bicluster with constant values, bicluster with coherent values on rows or columns and bicluster with coherent values.

Biclustering, block clustering, co-clustering and two-mode clustering are data mining techniques that allow synchronizing clustering of rows and columns of a gene expression matrix. Given a set of n rows in m columns (i.e., an $n \times m$ matrix), the co-clustering algorithm generates a co-cluster subset of rows that exhibit similar behavior across a subset of a column.

In addition to biclustering along the gene-sample dimensions, there has been a lot of interest in mining gene expression patterns across time. Hence, Triclustering finds the subset of genes that are similarly expressed across a subset of experimental conditions or samples over a subset of time points.

1.4. Triclustering

Given a time series gene expression data, a Triclustering algorithm aims to extract a set of triclusters such that each tricluster satisfy the properties like homogeneity and statistical significance. The homogeneity describes the structure, coherence and quality of a triclustering solution. The structure is conceived as the number, size, shape and position of triclusters. The coherence of a tricluster is defined by the observed correlation of values. The quality of a tricluster is the amount of tolerated noise. A tricluster is statistically significant if its occurrence probability deviates from expectations when it is unexpectedly low against a null data model. Fig. 2 shows Triclustering models (Henriques&Madeira2018).

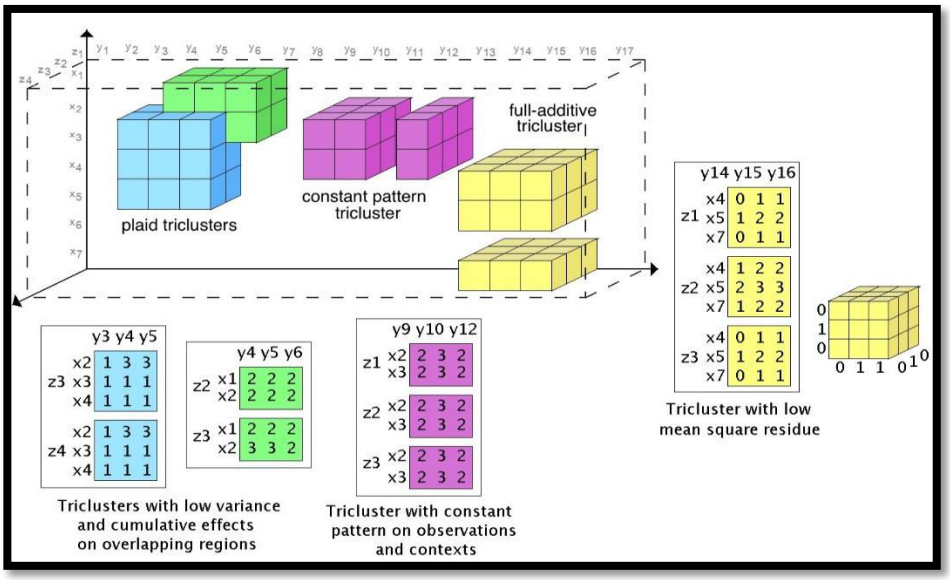

Fig. 2. A Triclustering Model

Triclustering algorithms aims to find genes that have similar expression profiles along a segment of timeseries in a subset of conditions. A coherent tricluster may contain a set of genes which exhibits either similar numeric values for the times and conditions (coherent values) or similar behaviors either as

correlated positive and negative changes in the expression values (coherent behavior) (Madeira & Oliveira2004). These types of coherent clusters contain information which helps in identifying useful phenotypes, potential genes that are related to these phenotypes and their regular relations (Tchagang et al. 2012).

In recent years, analyzing the microarray data for finding the coexpressed gene patterns has got the focus of researchers. Considering time series microarray datasets, there is a need for mining the gene expression patterns which are similarly expressed with both samples and time points. There are some existing triclustering algorithms for mining the triclusters from the gene expression dataset. This chapter deals with the review of existing triclustering algorithms for microarray datasets and also for other datasets.

2. Triclustering for Microarray Datasets

Mining coherent gene clusters from the Gene-Sample-Time (GST) point microarray dataset (Jiang et al. 2004) is proposed for extracting clusters that contain a subset of genes and samples that are coherent along with the time series. It performs gene-sample search and sample-gene search for completely extracting a coherent gene cluster. In sample-gene search, for each subset of sample combinations, a set of genes that are coherent are grouped and later some of the sample subsets could be pruned. Similarly, in gene-sample search, for each combination of genes, the set of samples that are coherent are grouped and later it could be pruned. Since the number of samples is lesser than the number of genes in the dataset, the sample-gene search performs better than the gene-sample search.

TriCluster (Zhao &Zaki2005) is the first introduced triclustering algorithm for extracting coherent clusters in three-dimensional datasets. It mines the overlapping and arbitrarily positioned clusters in which each cluster can have constant and similar values and also scaling and shifting patterns across each dimension. It constructs a graph-based model for mining the clusters. It splits each time point separately and develops a multigraph containing similar value ranges between two samples for each gene-sample matrix. It then finds the maximal cliques in that multigraph for extracting biclusters for that particular time point. Next, another multigraph is developed from the extracted biclusters of each time point and searching the cliques again from this graph will result in the set of triclusters. And, if needed the clusters can be merged or deleted for reducing the overlaps.

Sometimes, the absolute values of two dimension profiles in a graph may vary but their overall trend may remain consistent. According to biologists, the two genes are biologically associated when their trends are similar. TriCluster algorithm sticks to the symmetry property by considering only the absolute values of the genes, samples and time points. gTricluster (Jiang et al. 2006) algorithm breaks this property by extracting triclusters based on a general tricluster model. gTricluster applies a basic similarity metric called Spearman Rank Correlation for determining the similarity of two expression profiles. A new 3D cluster model is represented for GST microarray data. It also maintains the biological association of the genes with samples and time points. Initially the maximal coherent subset samples for each gene are identified and similarity matrix for each is computed. Among all the maximal cliques, depth first search is applied for extracting the triclusters.

Extended Dimension Iterative Signature Algorithm (EDISA) (Supper et al. 2007) is proposed for clustering 3D gene-condition-time datasets. Initial modules are constructed from the dataset which are refined by eliminating genes and samples until the members in the module compile with each other. Three possible modules are obtained such as module with independent response profiles, similar modules with similar responses under all conditions and a condition specific response. Pearson correlation is used for identifying similarity between two gene profiles.

Gene expression data of biological experiments have huge volume of data and in turn, it imposes burden in analyzing the data. In TriCluster algorithm, constructing multigraphs and combining the information to construct a graph again turns the triclustering problem to be NP complete which needs a parallelism concept to attain feasible solutions. ParTricluster algorithm (Araújo et al. 2008) is proposed for implementing the TriCluster algorithm in parallel using filter-labelled-stream in support with ant-hill parallel programming environment. The result analysis shows that it scales even well with large data size. It explores many parallelisms in filter stream models like task parallelism, data parallelism and asynchrony.

In the sight of biological analysis, to interpret gene regulatory mechanisms by using only gene expression data is hard to perform. Gene regulators information has the ability to control the gene expression value such as transcription factor binding site into the promoter region. So, in order to incorporate the gene regulator information with the gene expression data, Regulated Expression Values (REV) is introduced (Li & Tuck 2009). REV score indicates how a gene can be regulated by a factor. After assigning the REV scores, it directs the TriCluster algorithm to find the triclusters from the data with an addition feature such as Automatic Boundary Searching (ABS). It is added for determining the boundary threshold. It constructs the Transcription regulatory factor and Gene (T-G) matrix and Transcription factor and Conditions (T-C) matrix representing the regulatory information with genes and conditions. It extracts triclusters with each gene controlled by at least one regulator. But, constructing the REV score in high dimensional data is a very much complicated process.

LagMiner (Xu et al. 2009) is a 3D clustering algorithm with a cluster model S2D3 in which S2 gives the scaling-shifting correlation and D3 reflects three dimensions in the data genes, samples and time points. It aims at finding clusters in the gene-sample plane, time-gene plane and an order-preserving gene-time plane for each sample where it finds the time lag coefficient and time lag order. It achieves time-lag correlation, continuity of correlated time periods and general shifting and scaling patterns.

3D-TDAR-Mine (Liu et al. 2010) is proposed for mining the temporal dependency association rules in three-dimensional microarray data. The association rules that are mined using this algorithm provide the information on regulation of gene reactions. TS3 similarity measurement is introduced for finding the coherent pattern with shifting, scaling trends. TS3 uses the min-max normalization for normalizing the series which are in the same range for handling scaling and shifting. This algorithm undergoes two phases such as the coherent pattern phase for finding the coherent gene patterns and generating temporal association rules phase.

Versatile temporal subspace patterns (Hu &Bhatnagar 2011) discovery is proposed for analyzing the biological importance of three-dimensional data. A clear temporal pattern is extracted as a tricluster in searching for multiple contiguous subintervals. A prefix-based search algorithm is proposed with Pearson correlation method for finding the maximal triclusters from closed temporal patterns which may contain high meaningful biological hypotheses.

Intersected coexpressedsubcube miner (Ahmed et al. 2011) is proposed for mining both the inter-temporal and intra-temporal gene coherence in a dataset. It also eliminates the time-dominated, sample-dominated datasets and also detects the triclusters which are time latent. A method is proposed based on the order-preserving sub-matrices for mining the triclusters from three-dimensional data. From an unordered pair of gene-sample planes, a set of modules are generated which can be extended to form a tricluster. PMRS, Planar Similarity measure is also introduced for evaluating the triclusters.

Order Preserving Triclustering (OPT) (Tchagang et al. 2012) is a subspace clustering algorithm which is proposed for clustering only the three-dimensional short time series dataset. It uses the combinatorial approach on the sample dimension and applies the order preserving concept to the time dimension. These two approaches combinatorial and order-preserving allow finding the similarities and differences between samples in terms of temporal expression profiles. For the restriction of short time series data, the expression

profiles may vary from 2-5 samples and 3-8 time points. In the case of datasets that contain more time points and samples, the algorithm will work but it has a large computational complexity that increases with an increase in a number of time points or samples.

Three Way Clustering (TriWClustering) (Dede&Oğul 2013) algorithm is proposed for cross-species gene regulation analysis for mining then δ-triclusters from the dataset. This algorithm initiates with the entire dataset as a tricluster and then iteratively it removes rows, columns and sources of the tricluster until a residual score becomes less than or equal to δ. Next, it starts inserting rows, columns and sources until a stopping condition is met. Once a first tricluster is obtained, it starts the process again with the complete dataset.

δ-TRIMAX(Bhar et al. 2012) triclustering algorithm is proposed for extracting triclusters with a new fitness function MSR for the three-dimensional dataset. It is the extension of Cheng and Church biclustering algorithm that uses the MSR value for two-dimensional datasets. It follows a greedy heuristic technique for finding the triclusters but with a threshold δ. After extracting the triclusters, each one of the triclusters is represented by an Eigen value. Then, it finds whether the eigenvalue is expressed differentially at the early, middle or later Estrogen responsiveness stage. It also finds the hub genes which represent the triclusters, and binding site analysis is done for finding the transcription factor binding to a promoter region of the Estrogen receptor.

The triclustering plaid model 3D-Plaid (Mankad&Michailidis 2014) is introduced which is very similar to the existing plaid models but with an extension for the third dimension in the dataset. This algorithm finds subspaces which have strong deviations and finds their dependence with an iterative procedure aiming to minimize the sum of squares of the 3D data values. Initially, a background layer is fit and then the triclusters are then added one at a time until no more statistically significant triclusters can be found under a given permutation test.

TriGen algorithm (Aviles et al. 2014) is proposed for mining triclusters from temporal gene expression data by applying the optimization technique Genetic algorithm. An initial population representing genes, samples and time is generated. Then the quality of each individual is measured by a fitness function. Next, the individuals which would survive to the next generation are selected. Then, the genetic algorithm properties such as crossover and mutation are applied to each individual in the population. It is repeated until the stopping criterion is met. This algorithm is evaluated in terms of two measures namely, MSR and correlation measure.

Least Square Line (LSL) (Aviles&Escudero, 2014) measure is proposed for evaluating the quality of the tricluster. Three graphical views are considered such as one panel with X-axis as genes, Y-axis as samples, second panel with X-axis as genes and Y-axis as time points and third panel with X-axis as samples and Y-axis as time points. For evaluating the measure, Trigen algorithm is applied for extracting the triclusters. The results are also validated using two correlation measures such as Pearson and spearman correlation coefficients.

EMOA-δ-TRIMAX (Bharet al. 2015) is the Evolutionary Multi-Objective Optimization added to the δ-TRIMAX algorithm for retrieving the overlapping clusters by incorporating the features of the evolutionary algorithm. Genetic algorithm is applied for optimization by mutation for producing new offspring. Then, the population is ranked using dominance criteria and crowding distance as it replaces the worst population. Three objective functions are used for evaluating each individual in the population. The first objective function is MSR value divided by δ which is to be minimized; the second is the volume of the tricluster divided by the volume of the dataset which is to be maximized; the third is the non-parametric spearman correlation coefficient of the triclusters. It is also used for detecting the Eigen gene in δ-TRIMAX, singular valued decomposition on the expression data of each tricluster.

Three Way module Inference via Gibbs Sampling (TWIGS) (Amar et al. 2015) algorithm is proposed for finding the coherent and flexible modules in three-way data. It is based on the hierarchical Bayesian data

model and Gibbs sampling. It produces two modules in which the first module contains a set of rows that are active across a set of times and the second module contains a set of time points in each covered subject. It allows heterogeneity and asynchrony in different subjects in the first module and the second module discovers the specific regulatory information based on the dataset.

SSSimTri (Ahmed et al. 2014) is a Shifting and Scaling Similarity Triclustering algorithm for finding coexpressed patterns by shifting and scaling in gene-sample-time expression data. It uses a seed-growth algorithm in parallel for extracting biclusters from each time point's slice. A fast shared memory biclustering and shared nothing triclustering architecture is also proposed for identifying coexpressed patterns with higher biological significance over the gene-sample-time plane. This algorithm is also able to identify shifting, scalable and shifted-scalable patterns.

TimesVector (Jung et al. 2017) is a clustering algorithm for three-dimensional time series data for extracting clusters with distinctive gene expression patterns between more samples. Initially, dimension reduction and clustering of time-condition concatenated vectors are performed. Then, detecting similar and distinct gene expression patterns includes post-processing clusters. Finally, genes are rescued from unclassified clusters.

TriGen algorithm uses three fitness functions such as MSR, Least Square Lines(LSL) and Multi Slope Measure(MSL). In order to asses these results, a new evaluation measure is introduced called TRIQ (Avilés et al. 2018) is introduced. TRIQ acquires the information from three different sources such as correlation among gene, sample and times, graphical validation of the extracted patterns and function annotation of the extracted genes. BIOQ measures the biologically significant gene in the tricluster, GRQ gives the quantitative representation of the triclusters. PEQ and SPQ gives the pearson and spearman correlation of the genes in the triclusters respectively.

THD-Triclustering algorithm (Kakati et al. 2018) is developed for handling co-occurring shifting and scaling patterns from a dataset. This algorithm has two parts such as "generate biclusters" which extracts a set of biclusters that have high biological significance and "generate triclusters" which gives a set of triclusters having high biological significance and inter temporal coherence. This algorithm is validated using the coverage measure and biological significance analysis is done and thus key genes are identified.

Mean Correlation Value (MCV) (Narmadha&Rathipriya 2018) measure is introduced for identifying the correlation for the triclusters and also the triclusters are extracted from the dataset using this MCV. Its value ranges from 0 to 1 in which a value closer to 1 indicates highly correlated cluster. It extracts triclusters with additive pattern, multiplicative pattern, coherent pattern and coherent evolution pattern. The correlation based measure MCV is able to evaluate all types of the trilcusters since it can handle transformations like translation and scaling.

Particle Swarm Optimization technique is applied to extract the triclusters from the high dimensional data with objective function as Mean Square Residue (Swathypriyadharsin P & K Premalatha, 2019). The algorithm is applied to three real life microarray gene expression data which groups the coexpressed genes over a subset of samples under a subset of time points which imposes huge computational burden. The biological significances of the extracted triclusters from all the three datasets are also analyzed.

A novel rough fuzzy cuckoo search algorithm (Swathypriyadharsin P & K Premalatha, 2019) is proposed for triclustering genes across samples and time points simultaneously. By applying the upper and lower approximation of rough set theory and the objective function of fuzzy k-means, rough fuzzy k-means was incorporated into a cuckoo search to handle the uncertainty of the data. It was applied to three real-life time series gene expression datasets. This work was evaluated using four validation indices and correlation analysis was performed to indicate the cluster quality.

TriRNSC (BhawaniSankarBiswal et al., 2020) is a novel triclustering algorithm which is designed for the identification of significant triclusters within gene expression profiles. TriRNSC builds upon the restricted neighborhood search clustering (RNSC) method, a widely utilized graph-based clustering approach that

takes into account genes, experimental conditions, and time points simultaneously. TriRNSC framework initially employs a basic cost function before transitioning to a scaled cost function for the analysis of gene expression. Furthermore, TriRNSC gets impressive results by incorporating gene size as an evaluative parameter for the generated triclusters.

δ-Trimax (Siswantining et al., 2021) method is employed for triclustering analysis on microarray gene expression data which aims to identify triclusters with a mean square residual smaller than δ and maximum volume. The triclusters are derived by selectively removing nodes from 3D data using both multiple node deletion and single node deletion algorithms. To refine the obtained tricluster candidates, a node addition algorithm is employed, reintroducing previously deleted nodes for further evaluation. Improvements to the δ-Trimax method are implemented, and the evaluation of the resulting triclusters is also calculated. The method is applied to two microarray gene expression datasets. The first dataset involves gene expression data from the differentiation process of human-induced pluripotent stem cells (HiPSCs) from patients with heart disease. The best simulation is achieved when δ=0.0068, λ=1.2, resulting in five triclusters considered characteristic of heart disease. The second implementation is on HIV-1 data, with the optimal simulation parameters being δ=0.0046 and λ=1.25. This yields three genes, namely AGFG1, EGR1, and HLA-C, identified as biomarkers. This gene set holds potential for medical experts in guiding further treatment strategies. The best five tricluster based on the smallest TQI for HiPSC data provides group of gene expression within the five tricluster is supposed to be a feature of heart disease. Therefore, this gene group can be used by medical experts in providing further treatment, such as making the genes in this tricluster a therapeutic target or as a drug development. Three biomarkers for HIV-1 disease were obtained from the 10 selected tricluster. Biomarkers consist of genes AGFG1, EGR1, and HLA-C.

3. Triclustering for other Social Three Dimensional Datasets

Mining 3D Subspace Clusters (MIC) (Sim et al. 2010) is proposed for extracting triclusters from 3D datasets with constant coherence. The triclusters extracted are called Correlated 3D Subspace Clusters. It has a high correlation within the context slices and between each pair of contiguous slices. Initially, it generates seeds that are correlated objects and later extends it by merging the seeds that maximize correlation score.

Triclustering is also applied for frame induction problem that involves the generalization of clustering for triadic data. The process of frame net construction is automatically done through unsupervised learning techniques. Triframes (Ustalov et al. 2010), a graph-based approach is proposed as it provides state-of-the-art results on frame net derived datasets and also performs the verb class clustering task. Here, the focus is on the subject-verb object triples and two frame roles that are expressed by subjects and objects which gives the extracted semantic structures with high coverage.

Tribox cluster (Mirkin&Kramarenko 2011) is a disjunctive model for clustering the binary data using least squares locally optimal method. It involves choosing the right parameter, scale shift and produces a contrast box of triclusters. It selects triples from the data and modifies its extent, intent and modus by maintaining high density. TriBox extracts triclusters by maximizing an objective function. Hash functions of the corresponding triclusters may also be used for optimizations.

The incorporation of domain knowledge in the clustering process is not done by most of the existing algorithms and many such algorithms are dependent on the parameters in which setting the wrong threshold may reduce the quality of the clusters. To address these issues, a centroid-based actionable 3D subspace clustering algorithm called CATSeeker (Sim et al. 2013) is introduced which allows the inclusion of the domain knowledge information and becomes parameter insensitive. This algorithm contains a unique

combination of singular value decomposition, numerical optimization and three-dimensional frequent itemset mining.

One pass algorithm (Gnatyshak et al. 2014) is the extension of OAC triclustering which is designed for binary data having linear time and memory complexities. Next, this algorithm is parallelised using the map-reduce framework for big data. There exist two stages M/R approach. In the first stage, M/R calculates all the existing pairs of primes efficiently. The second stage of M/R permits for assembling of the found primes into triclusters. The number of map keys is equal to a number of reducers.

Two approaches are developed for triclustering the binary data. The first approach citation considers tricluster as a subset of ternary relation Y that contains objects, attributes and conditions. The second approach citation is to find a dense sub-matrix of the adjacency matrix of ternary relation Y. Hierarchical spectral triclustering (Ignato vet al. 2014) is proposed for extracting the dense submatrices of the adjacency matrix of the initial ternary relation Y. Traditional spectral partitioning is the reduction of bipartite graph and this method is extended to the tripartite graphs. It has good scalability on real-world data and is the other alternative for conceptual triclustering.

Triadic formal concept analysis (Ignatovet al. 2015) is done for optimal patterns in which a dense maximal cuboids called triadic pattern is constructed. Triclusters that are optimal with respect to the least-square criterion and graph partitions are obtained with the help of spectral clustering. Finding an optimal tricluster is a NP complete problem which needs extensive computational experiments that are done applying pareto-optimality principle. The evaluation of the algorithm is based on five criteria such as density, coverage, diversity, noise tolerance and cardinality. Choosing the right number of clusters is an issue which needs to be addressed.

A single pass triclustering algorithm (Gnatyshak 2015) for extracting the clusters is proposed for the big data that clusters the Object, Attributes and Conditions (OAC). It is an OAC triclustering method but it is based on the three prime attributes such as prime object attribute, prime object condition and prime attribute condition. Thereby it enumerates all the triples of three relationships and a tricluster is generated for each set. Then for each tricluster, the hash value is computed and the corresponding tricluster is added to the hash table. If any conflicts occur, the tricluster should not be added to the hash table.

OAC triclustering algorithm (Zudin et al. 2015) is applicable only to the binary data, so it is extended to work for other data by modifying the basic OAC triclustering approach. The algorithm is parallelised using the map-reduce framework for adapting the algorithm for the big datasets. It has linear time and reduces memory complexities.

Traditional Fuzzy clustering allows an object to be presented in multiple clusters without being restricted to a single cluster. Fuzzy Co-clustering extends the fuzzy clustering by calculating membership functions for both the objects and features. Further, it is extended for clustering the three-dimensional data by proposing a Fuzzy Triclustering (FTC) algorithm (Liu et al. 2015). For each dimension, a fuzzy membership function is assigned by the FTC algorithm and it extracts the clusters simultaneously on three dimensions such that the members inside a cluster have strong coherence with each other. The algorithm is evaluated using two measures such as precision and recall. However, determining the number of clusters is a major issue in FTC.

Triclustering approach (Guigoures et al. 2012) is also applied for tracking the structures in time-evolving graphs in which three-dimensional coclustering of source vertices, target vertices and time is considered as three dimensions. It allows inferring the time segments from the evolution of edge distribution between vertices that does not require the user to specify a priori discretization.

Most of the clustering methods are for either temporal or spatial data but there also exists large and complex spatio-temporal datasets. So, Bregman Cuboid Average Tri-Clustering (BCAT) (Milla et al. 2018) is proposed for analyzing the data cubes of such data. The three dimensions considered here are space, time and a nested temporal or spatial dimension. It identifies the triclusters with members having similar

temperature values along with consideration of spatial weather stations. Next, the k-means algorithm is used for mining the number of irregular triclusters.

G-Tric (João Lobo et al., 2021) aims to create synthetic datasets with configurable properties and the possibility to plant triclusters. The generator is prepared to create datasets resembling real 3-way data from biomedical and social data domains including triclustering solution as output. G-Tric can replicate real-world datasets and create new ones that match the researcher's needs across several properties, including data type (numeric or symbolic), dimensions, and background distribution. Users can tune the patterns and structures that characterize the planted triclusters (subspaces) and how they interact (overlapping). Data quality can also be controlled, by defining the amount of missing, noise or errors. Triclustering evaluation using G-Tric provides the possibility to combine both intrinsic and extrinsic metrics to compare solutions that produce more reliable analyses. A set of predefined datasets, mimicking widely used three-way data and exploring crucial properties was generated and made available, highlighting G-Tric's potential to advance triclustering state-of-the-art by easing the process of evaluating the quality of new triclustering approaches. Besides the ability to easily generate customized three-way data with triclustering solutions, it also enables the possibility to perform benchmarks on existing algorithms to study their efficiency within certain conditions, or their effectiveness in finding different types of patterns, by allowing the creation of several datasets with an extensive board of characteristics.

A bigTriGentriclustering algorithm based on evolutionary algorithms (Laura Melgar-García et al., 2022) is introduced to mine the three-dimensional patterns on the basis of vegetation indices from vine crops. Different vegetation indices have been tested to find different patterns in the crops. The bigTriGen algorithm has been applied to a vineyard crop in southern Portugal for finding a precision viticulture solution. The accuracy of the algorithm has been shown with respect to two different features: the quality measure of the found patterns and the scalability of the algorithm. Different vegetation indices have been calculated using Sentinel-2 images downloaded from QGIS software and it was found that the patterns using these vegetation indices have shown that the index that best fits this field is the MSI. In this way, the algorithm has been able to find four different areas of the vineyard crop that behave differently in terms of their soil moisture. The scalability of the algorithm has been studied considering the number of nodes used and the size of the dataset.

TCtriCluster (Diogo F. Soares et al., 2023) is a temporally constrained triclustering algorithm that aims to systematically identify informative temporal patterns shared by a subgroup of patients in specific features (triclusters). These identified patterns serve as discriminative features in a cutting-edge classifier, ensuring interpretability. This methodology provides model interpretability by uncovering clinically relevant disease progression patterns, and elucidating features crucial for classification. To assess the prognostic boundaries of five significant clinical endpoints—non-invasive ventilation (NIV) requirement, auxiliary communication device necessity, percutaneous endoscopic gastrostomy (PEG) need, caregiver requirement, and wheelchair necessity—the TCtriCluster approach is applied to the Amyotrophic Lateral Sclerosis (ALS) Portuguese cohort (N = 1321). The results reveal that triclustering-based predictors outperform existing state-of-the-art alternatives. Notably, predictions for the need for an auxiliary communication device (within 180 days) and the requirement for PEG (within 90 days) achieve an AUC above 90%. Validated in clinical practice, this approach aids healthcare professionals in comprehending the intricate and varied patterns of ALS disease progression, thereby contributing valuable insights into prognosis. The possibility of extracting group-specific patterns along time frames of arbitrary length offers a higher degree of feature expressiveness, which is generally lacking in peer approaches.

TriSig (Leonardo Alexandre et al., 2023) is a novel triclustering methodology to rigorously assess the statistical significance of patterns extracted from tensor data. The methodology provides a robust set of statistical principles that accommodate different aspects of tensor data such as variable domains and dependencies, temporal dependencies and misalignments, and relevant p-value corrections. This work aims

at proposing a statistical frame to assess the probability of patterns in tensor data to deviate from null expectations, extending well-established principles for assessing the statistical significance of patterns in matrix data. A comprehensive discussion on binomial testing for false positive discoveries is achieved through variable dependencies, temporal dependencies and misalignments, and p-value corrections under the Benjamini-Hochberg procedure.

4. Review Analysis

Table 1 provides a detailed review report of all the triclustering algorithms that have been discussed so far. All the algorithms were compared in terms of the approach it used, the measure it followed, the dataset employed and the validation framework applied. It also depicts the evolution of the triclustering of three-dimensional data over the years, starting from a simple pattern-based approach it then extended to graph-based, greedy divide and conquer, stochastic and exhaustive and finally arrived at the applications of evolutionary optimization approaches in dealing with the three-dimensional data. The most commonly used measures are the MSR and correlation scores. The datasets applied by different algorithms are also provided in the table which are all three-dimensional data including the gene expression data. The validation framework used in most of the gene expression data is biological significance analysis of the extracted genes by the algorithm.

Table 1 Review of Triclustering Algorithms

Algorithm	Year	Approach	Measure	Dataset	Validation Framework
Mining Coherent clusters from GST	2004	Pattern based	Pearson Correlation	GST microarray data of multiple sclerosis	Gene ontology and p-value
TriCluster	2005	Graph based	Coverage, Overlap	Yeast Cell Cycle dataset	Gene Ontology and p-value
gTricluster	2006	Pattern based	Spearman rank correlation	Yeast cell cycle	Precision and Recall and Gene ontology
EDISA	2007	Greedy (seed growth and reduction)	Pearson Correlation	Homo sapiens and Arabidopsis thaliana root and Arabidopsis thaliana shoot	Biological significance analysis
ParTricluster	2008	Graph based with anthill environment		Yeast cell cycle regulated dataset	Edge traversal and computational cost
TRIClustering	2009	Automatic Boundary Searching	Regulation coefficient and similarity score	Yeast Sporulation REV data	Regulation Profile and TRANSFAC
LagMiner	2009	Greedy (seed growth with multi-wise intra-plane coherence)	Shifts/scales on X and Z slices; order-preserving on Y	One synthetic dataset and one real-life yeast 3D dataset	Sensitivity, Effectiveness and yeast genome ontology
MIC	2010	Greedy (seed growth)	3D correlation information	Stock Market Data	Run time and cluster quality
Triframes	2010	Graph based approach	SVO triples	FrameNet derived dataset	Fscore for SVO and ploysemous verb classes
3D-TDAR-Mine	2010	Pattern-based (quasi-exhaustive)	Associative subspaces with shifts, scales and trends on X	Yeast Saccharomyces cerevisiae	Coherence Threshold Analysis, Rule Analysis, Similarity Factors Analysis, Rule Redundant Rate Analysis
Intersected coexpressedsubcube miner	2011	Greedy approach	Intra temporal homogeneity and Inter temporal homogeneity	Saccharomyces Cell Cycle and two Yeast Cell Cycle	Planar Similarity measure
Versatile temporal subspace patterns	2011	Pattern-based (quasi-exhaustive)	LOESS-based Pearson on X slices	One synthetic dataset and GSE20635 genomic dataset from GEO	Gene Ontology
Time Evolving	2012	Stochastic approach		London cycles,	Mutual Information

Algorithm	Year	Approach	Measure	Dataset	Validation Framework
Graphs				optimal Image and Simplified Image	
OPTricluster	2012	Combinatorial approach with order preservation		Mice, Arabidopsisthaliana, Brassicanapus	Gene Ontology and p-value
OPT	2012	Combinatorial approach		Plasmodium chabaudi, Arabidopsis thaliana, Brassica napus and Brassica napus	Statistical significance and complexity analysis
CATSeeker	2013	Singular value decomposition and 3D frequent itemsetmining		synthetic, proteinstructuralandfinancial data	Efficiency Analysis, parametersensitivityanalysis,quality analysis of theclusters mined by differentalgorithms,applicationon stock market, and applicationon protein structure
δ-TRIMAX	2013	Greedy (Divide and Conquer)	Mean Square Residue	HomoSapiensestrogen induced breast cancer dataset	GO and Kegg pathway enrichment analysis and TFBS enrichment analysis
TriWClustering	2013	Greedy (divide and conquer)	Fully-additive (3D MSR)	3 GEO datasets	Gene Ontology term enrichment analysis and Dunn index (DI) metric
TriGen	2014	Optimization based	MSR and LSL	Synthetic dataset and Sachharomyces cerevisiae	Coverage, overlapping and Gene ontology, p-value
SpecTric	2014	greedy (divide and conquer)		Bibsonomy and Amsterdam-Amstelland police report	Density, diversity and coverage
3DPlaid	2014	Stochastic	Plaid assumption	T-cell data and World Trade data	RunTime and Computational cost
EMOA- δ-TRIMAX	2015	Optimization based	MSR	Two synthetic datasets and Three real datasets from GEO as GSE11324,	SDB, TQI, GO and Kegg pathway enrichment analysis and TFBS enrichment analysis

Algorithm	Year	Approach	Measure	Dataset	Validation Framework
				GSE35671, GSE46280	
Fuzzy TriCluster	2015	Fuzzy		MovieLens Dataset	F-measure, entropy, overall Similarity
TWIGS	2015	Hierarchical Bayesian data model and Gibbs sampling		Microarray datasetfortranscriptionalresponse of patientsto sepsis and fMRIdata	p-value
Triadic formal concept analysis	2015	Exhaustive		Mobile operators, movies and bibsonomy	Density, coverage, diversity, noise tolerance and cardinality
Single pass Triclustering algorithm	2015	Exhaustive		Two synthetic datasets and IMDB and Bibsonomy	Density and coverage
OAC triclustering wit Map Reduce	2015	Exhaustive		Two synthetic datasets and IMDB and Bibsonomy	Execution time
SSSimTri	2016	Parallelized biclustering-based	Shifts and/or scales on Z slices	Yeast Sporulation, Yeast Cell Cycle, Rat CNS and Homo sapiens	Gene Ontology and p-value
TimesVector	2017	Pattern-based	Similar and differential patterns (cosine distance)	Four Microarray datasets	Biological significance analysis
THD-Tricluster	2018	Similarity based	Shifting-and-Scaling Similarity (SSSim)	HIV-1 progression data	Coverage, biological process and p-value
Particle Swarm Optimization based Triclustering	2019	Optimization	Mean Square Residue, SDB & Average TQI	Three real-life time series gene expression datasets of Homosapiens, Mus Musculus& Saccharomyces cerevisiae	Biological Significance analysis
Rough fuzzy cuckoo search algorithm	2019	Optimization	Mean Square Residue & Coverage, Average	Three real-life time series gene expression	Biological Significance analysis

Algorithm	Year	Approach	Measure	Dataset	Validation Framework
			row variance & Average Correlation value	datasets of Homosapiens& two Mus Musculus	
TriRNSC	2020	graph-based clustering approach	TQI, SDB & MSR	Yeast cell data set	Gene ontology &Kegg pathway analysis
δ-Trimax	2021	Greedy (Divide and Conquer)	Mean Square Residual, TQI	HiPSCs& HIV-1 dataset	
G-Tric	2021	Pattern based approach	Missing, Noise & Errors	Yeast cell cycle, Stock market ratios, fMRI-Average blood-oxygen-level-dependent contrast, Bibsonomy, Georeferenced time-series	
bigTriGen	2022	Evolutionary algorithmic approach	Moisture Stress Index (MSI) and the Green Normalized Difference Vegetation Index (GNDV I)	Vineyard crop experimental dataset	Scalability analysis
TriSig	2023	Pattern discovery	p-value	Tensor data	Pattern evaluation
TCtriCluster	2023	Stochastic	AUC	Lisbon ALS clinic dataset	

5. Summary

For mining the useful information from three-dimensional datasets, many Triclustering algorithms were proposed. It has a computational challenge in microarray data analysis due to its three-dimensional characteristics. This paper provides a detailed study of the different existing triclustering approaches developed for microarray gene expression data, binary data and big data. Three-dimensional data have complexity in dealing with the third dimension of the data, so many triclustering algorithms were applied to extract the meaningful similar properties in all three dimensions of the data. In recent years, many new triclustering algorithms are proposed but there is no single best algorithm that is successful in all aspects. Heuristic optimization techniques are also applied for triclustering the three-dimensional time series datasets. Triclustering algorithms are divided into two main categories such as applying triclustering techniques on the microarray gene expression data and the other one is mining the useful hidden patterns from other kinds of three-dimensional data. The paper discusses the different types of approaches that have been followed in triclustering algorithms such as pattern-based, graph-based, combinatorial, stochastic, greedy divide and conquer, heuristic optimization and evolutionary approach. The main measure used in

most of the triclustering algorithms is MSR (Mean Square Residue). The most common validation framework used is the coverage and similarity index. In the gene expression datasets, gene ontology, enrichment analysis and biological significance analysis are used as the validation framework.

The review concludes by insisting the challenges still need to addressed in the area of triclustering three-dimensional data like (1) the scalability of the triclustering algorithm needs to be improved as the increase in dataset size might have an impact on the complexity overhead in dealing with three dimensions, (2) statistical analysis can be included in order to evaluate the significance of the triclusters, (3) developing an integrative framework for combining the dispersed potentialities of the existing best algorithms, (4) evolving with different strategies for dealing the temporal misalignments of data, (5) Enhancing the algorithms by including the degree of tolerance to different kinds of noise in a huge three-dimensional data.

Statements and Declarations

Competing Interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding Information

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Author contribution

Dr. P. Swathypriyadharsini confirms responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Dr. K. Premalatha confirms responsibility for the following: Supervision, Project Administration, Writing - Review & Editing

Data Availability Statement

The data used in this research is available at Whitehead Institute Center for Genomic Research: cancer genomics www-genome.wi.mit.edu.

Research Involving Human and /or Animals

This research study does not involve the use of animals and humans including for any testing, experimentation, or participation.

Informed Consent

Informed consent is not required for this study as it adheres to the ethical standards and regulations.

References:

1. Jiang, D, Tang, C & Zhang, A 2004, 'Cluster Analysis for Gene Expression Data: A Survey', IEEE Transactions on Knowledge & Data Engineering, vol. 16, no. 11, pp. 1370-1386.
2. Pollard, KS & Van Der Laan, MJ 2002, 'Statistical Inference for Simultaneous Clustering of Gene Expression Data', Mathematical Biosciences, vol. 176, no. 1, pp. 99-121.
3. Cheng, Y& Church, GM 2000, 'Biclustering of expression data', In Proceedings of International Conference on Intelligent Systems Molecular Biology, pp. 93-103.
4. Henriques, R & Madeira, CS 2018, 'Triclustering Algorithms for Three-Dimensional Data Analysis: A Comprehensive Survey', ACM Computing Surveys, vol. 51, no. 5, pp. 1-43.
5. Madeira, SC & Oliveira, AL 2004, 'Biclustering algorithms for biological data analysis: A survey', IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 1, no. 1, pp. 24-45.

6. Tchagang, AB, Phan, S, Famili, F, Shearer, H, Fobert, P & Huang, Y 2012, 'Mining biological information from 3D short time-series gene expression data: the OPTricluster algorithm', Bioinformatics, vol. 13, no. 1, pp. 54.
7. Jiang, D, Pei, J, Ramanathan, M, Tang, C & Zhang, A 2004, 'Mining coherent gene clusters from gene-sample-time microarray data', Proceedings in KDD Conference.
8. Zhao, L &Zaki, MJ 2005, 'TRICLUSTER: an effective algorithm for mining coherent clusters in 3D microarray data', Proceedings of the 2005 ACM SIGMOD international conference on Management of data, pp. 694-705.
9. Jiang, H, Zhou, S, Guan, J & Zheng, Y 2006, 'gTRICLUSTER: A More General and Effective 3D Clustering Algorithm for Gene-Sample-Time Microarray Data', Data Mining for Biomedical Applications, vol. 3916, pp. 48–59.
10. Supper, J, Strauch, M, Wanke, D, Harter, K, & Zell, A 2007, 'EDISA: extracting biclusters from multiple time-series of gene expression profiles', BMC Bioinformatics, vol. 8, pp. 334.
11. Araújo, RB, Ferreira, GHT, Orair, GH, MeiraJrW, Ferreira, RAC, Neto, DOG &Zaki, MJ 2008, 'The ParTriCluster Algorithm for Gene Expression Analysis', International Journal of Parallel Programming, vol. 36, no. 2, pp. 226-249.
12. Li, A, & Tuck, D 2009, 'An Effective Tri-Clustering Algorithm Combining Expression Data with Gene Regulation Information', Gene Regulation and Systems Biology, vol. 3, pp. 49-64.
13. Xu, X, Lu, Y, Tan, K & Tung, AKG 2009, 'Finding Time-Lagged 3D Clusters', 2009 IEEE 25th International Conference on Data Engineering, pp. 445-456.
14. Liu, Y, Lee, C, Chen, W, Shin, JW, Hsu, H & Tseng, VS 2010, 'A novel method for mining temporally dependent association rules in three-dimensional microarray datasets', 2010 International Computer Symposium (ICS2010), Tainan, pp. 759-764.
15. Hu, Z &Bhatnagar, R 2011, 'Discovery of Versatile Temporal Subspace Patterns in 3-D Datasets', 2011 IEEE 11th International Conference on Data Mining, Vancouver,BC, pp. 1092-1097.
16. Ahmed, HA, Mahanta, P, Bhattacharyya, DK, Kalita, JK & Ghosh, A 2011, 'Intersected coexpressedsubcube miner: An effective triclustering algorithm', 2011 World Congress on Information and Communication Technologies, Mumbai, pp. 846-851.
17. Dede, D &Oğul, H 2013, 'A three-way clustering approach to cross-species gene regulation analysis', 2013 IEEE INISTA, Albena, pp. 1-5.
18. Bhar, A, Haubrock, M, Mukhopadhyay, A, Maulik, U, Bandyopadhyay, S &Wingender, E 2012, 'δ-TRIMAX: Extracting triclusters and analyzing coregulation in time series gene expression data', Springer, vol. 7534, pp. 165-177.
19. Mankad, S &Michailidis, G 2014, 'Biclustering Three-Dimensional Data Arrays With Plaid Models, Journal of Computational and Graphical Statistics, vol. 23, no. 4, pp. 943-965.
20. Aviles, DG, Escudero, CR, Alvarez, FM &Riquelme, JC 2014, 'TriGen: A genetic algorithm to mine triclusters in temporal gene expressison data', Neurocomputing, vol. 132, pp. 42-53.
21. Avilés, GD&Escudero, RC 2014, 'LSL : A new measure to evaluate triclusters', IEEE International Conference on Bioinformatics and Biomedicine, pp. 30-37.
22. Bhar, A, Haubrock, M, Mukhopadhyay, A, Maulik, U, Bandyopadhyay, S &Wingender, E 2015, 'Multiobjectivetriclustering of time-series transcriptome data reveals key genes of biological processes', BMC Bioinformatics, vol. 16, pp.200.
23. Amar, D, Yekutieli, D, Katz, AM, Hendler, T, & Shamir, R 2015, 'A hierarchical Bayesian model for flexible module discovery in three-way time-series data', Bioinformatics, vol. 31, no. 12, pp. 17-26.

24. Ahmed, HA, Mahanta, P, Bhattacharyya, DK &Kalita, JK 2014, 'Shifting-and-Scaling Correlation Based Biclustering Algorithm', IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 11, no. 6, pp. 1239-1252.
25. Jung, I, Jo, K, Kang, H, Ahn, H, Yu, Y, & Kim, S. 2017, 'TimesVector: a vectorized clustering approach to the analysis of time series transcriptome data from multiple phenotypes', Bioinformatics, vol. 33, no. 23, pp. 3827-3835.
26. Avilés, DG, Giráldez, R, Cumbreras, FJG &Escudero, RC 2018, 'TRIQ: a new method to evaluate triclusters', BioData Mining, vol. 11, pp. 15.
27. Kakati, T, Ahmed, HA, Bhattacharyya, DK, &Kalita, JK 2018, 'THD-Tricluster: A robust triclustering technique and its application in condition specific change analysis in HIV-1 progression data', Computational biology and chemistry, vol. 75, pp. 154-167.
28. Narmadha, N &Rathipriya R 2018, 'Triclustering Algorithm for 3D Gene Expression Data using Correlation Measure', International Journal of Engineering Research in Computer Science and Engineering , vol. 5, no. 2, pp. 221-227.
29. P. Swathypriyadharsini, & K. Premalatha 2019, "Particle Swarm Optimization for Triclustering High Dimensional Microarray Gene Expression Data", Research Journal of Pharmacy and Technology, vol. 12, no. 5, pp. 2222-2228.
30. P. Swathypriyadharsini, & K. Premalatha 2019, "Rough Fuzzy Cuckoo Search for Triclustering Microarray Gene Expression Data" Turkish Journal of Electrical Engineering and Computer Sciences, vol. 27, no. 6, pp. 4328-4339.
31. Biswal, Bhawani, Patra, Sabyasachi, Mohapatra, Anjali &Vipsita, Swati 2020, "TriRNSC: triclustering of gene expression microarray data using restricted neighbourhood search", IET Systems Biology, vol. 14, no. 6, pp. 323-333.
32. Siswantining, Titin, NovalSaputra, DevviSarwinda, and HerleyShaori Al-Ash. 2021, "Triclustering Discovery Using the δ-Trimax Method on Microarray Gene Expression Data" Symmetry, vol. 13, no. 3: 437.
33. Sim, K, Aung, Z &Gopalkrishnan, V 2010, 'Discovering Correlated Subspace Clusters in 3D Continuous-Valued Data', Proceedings - IEEE International Conference on Data Mining, ICDM, pp. 471-480.
34. Ustalov, D, Panchenko, A, Kutuzov, A, Biemann, C &Ponzetto, SP 2018, 'Unsupervised Semantic Frame Induction using Triclustering', CoRR, 2018.
35. Mirkin, BG &Kramarenko, AV 2011, 'Approximate bicluster and tricluster boxes in the analysis of binary data', RSFDGrC'11, Moscow, Russia Berlin, Heidelberg: Springer-Verlag, pp. 248–256.
36. Sim, K, Yap, G, Hardoon, DR, Gopalkrishnan, V, Cong, G &Lukman, S 2013, 'Centroid-Based Actionable 3D Subspace Clustering', IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 6,
37. pp. 1213-1226.
38. Gnatyshak, D, Ignatov, DI, Kuznetsov, SO, &Nourine, L 2014, 'A One-pass Triclustering Approach: Is There any Room for Big Data?', CLA.
39. Ignatov, DI, Kuznetsov, SO, Poelmans, J, & Zhukov, L. 2013. 'Can triconcepts become triclusters?', Int. J. General Systems, vol. 42, pp. 572-593.
40. Ignatov, DI, Gnatyshak, DV, Kuznetsov, SO & Boris GM 2015, 'Triadic Formal Concept Analysis and triclustering: searching for optimal patterns', Mach Learn, vol. 101, no. 1, pp. 271-302.
41. Gnatyshak, D2015, 'A single-pass triclustering algorithm', Automatic Documentation and Mathematical Linguistics, vol. 49, pp. 27-41.
42. Zudin, S, Gnatyshak, D&Ignatov, DI 2015, 'Putting OAC-triclustering on MapReduce', CLA.

43. Liu, Y, Yang, T & Fu, L 2015, 'A Partitioning Based Algorithm to Fuzzy Tricluster', Mathematical Problems in Engineering, vol. 2015, pp.1-10.
44. Guigoures, R, Boulle, M &Rossi, F 2012, 'A Triclustering Approach for Time Evolving Graphs', Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops, IEEE Computer Society, Washington, DC, USA, pp. 115-122.
45. Milla, RZ, Wu, X, Verdiguier, EI &Kraak, MJ 2017, 'Tri-clustering : a novel approach to explore spatio-temporal data cubes', Proceedings International Conference on GeoComputation, pp. 1-5.
46. Lobo, J., Henriques, R. & Madeira, S.C. 2021, "G-Tric: generating three-way synthetic datasets with triclustering solutions", BMC Bioinformatics, vol. 22, no. 16.
47. Melgar-García, Laura & Gutiérrez-Avilés, David &Godinho, Maria Teresa &Espada, Rita & Brito, Isabel &Martínez-Álvarez, Francisco &Troncoso, Alicia & Rubio-Escudero, Cristina. 2022, "A new big data triclustering approach for extracting three-dimensional patterns in precision agriculture", Neurocomputing, vol. 500, no. 3.
48. Soares DF, Henriques R, Gromicho M, de Carvalho M, Madeira SC. 2023, "Triclustering-based classification of longitudinal data for prognostic prediction: targeting relevant clinical endpoints in amyotrophic lateral sclerosis", Sci Rep., vol. 3, no. 1, pp. 6182.
49. Alexandre, Leonardo & Costa, Rafael &Henriques, Rui. 2023, "TriSig: Assessing the statistical significance of triclusters", Pattern Rocognition.