Application Of Clustering Data Mining Techniques In Temporal Data Sets Of Hydrology: A Review

Rakesh Purviya[1], H.L. Tiwari[2] and Satanand Mishra[2]
[1]PG Scholar, Dept. of Civil Engg., MANIT, Bhopal,[2]Asst. Professor, Dept. of Civil Engg., MANIT, Bhopal, [2]Sciencist, CSIR-AMPRI, Bhopal.
Corresponding Email: [1]purviya88rakesh@gmail.com, [2]hltiwari@rediffmail.com, [2]snmishra07@gmail.com

Abstract: Hydrologic cycle are rather very complex and it is very difficult to predict the behaviour of runoff based on temporal data sets of hydrological process, as these are often very large and difficult to analyse and display. Clustering can be done by the different number of algorithms such as hierarchical, partitioning, grid and density based algorithms. This paper is original concerns in two main aspects. First, it provides an evolutionary algorithm for clustering starting from data mining mechanism, tasks and its learning. Second, it provides a taxonomy that highlights some very important aspects in the context of clustering algorithms, namely, hierarchical, partitional algorithms, density based, grid based and model-based. A number of references are provided that describe applications of evolutionary algorithms for clustering in different domains as well as in Hydrology. Also, in this paper a brief overview of temporal data mining concepts including time series sequences are discussed.

Keywords: *Temporal, Clustering, Data mining, Hierarchical, Hard and soft clustering, Hydrological process, Time series sequences.*

I. Introduction

A hydrologic process is a phenomenon describing the occurrence and movement of water in the earth phase of the hydrologic cycle. Developing a hydrological model based on past records is crucial and effective in many water resources applications such as optimal reservoir operation, drought management, flood control, hydropower generation and sustainable development of watershed area, etc. For many hydrological problems, sample data is sometime being very large and uncontrolled. Moreover the collected data involved some hidden sources of error. To handle such vast multiple variable data we need a technique that sorts such data [35]. Data mining is that branch of computer science, which is capable for extraction of valuable information hidden in the datasets for a given hydrologic process. Data mining can be defined as an activity or a process that extracts some new nontrivial information contained in large databases [21]. Data mining involves the oddment or rarity detection, association rule learning, classification, regression, summarization and clustering [4]. Due to blistering increase in storage of data, the stake in the discovery of hidden information in databases has exploded in the last decade. It is something like a big bang explosion in databases. Particularly, the clustering of time series has attracted the interest of researchers.

Cluster Analysis is an automatic process to find out similar objects from a given database. As noteworthy, it is one of the fundamental operations in data mining [4]. Fig 1 shows the complete clustering procedure involved in a given temporal data sets [6].

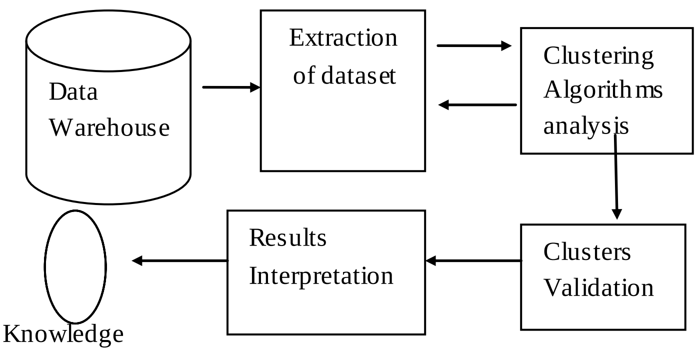

Fig. 1 Clustering procedure [6]

Clustering is a process of grouping objects with similar properties. A cluster is an assemblage of data objects that are similar to one another within the same cluster and dissimilar to the objects in other cluster. In data mining the data is mined using two learning approaches i.e. supervised or unsupervised learning (clustering). Clustering is an unsupervised learning i.e. it learns by observation rather than examples as there are no predefined class label exists for the data points in clustering process [7]. If there exists any predefined class then this will be regarded as Classification. Main task of clustering are explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics[8] [9] [10].

In this text, an attempt has been developed to provide comprehensive coverage of clustering techniques and their application in current engineering trends. The design of this paper is presented in such a way that it describes associated clustering techniques of data mining processes for developing Rainfall-Runoff Models. The remainder of the paper is organised as follows. Section 2 gives an overview of temporal data mining and along with different types of temporal data. Discussion further proceeds to Section 3, which gives a brief literature survey on applications of clustering techniques and its various algorithms, not only in Hydrology but also in various merging trends. Section 4 describes different clustering algorithms viz. Hierarchical clustering algorithms, K-means clustering algorithms, and Density Based Clustering Algorithm and also, the parameter used in these algorithms are described. Finally in section 5, the conclusions and proposed work are provided.

II. Temporal Data Mining

In this section, we will first review the concepts of temporal data mining and how is differ from conventional time series sequences is depicted, then its various tasks along with different classes are described.

Temporal data mining is concerned with extraction of hidden information of large sequential data sets. Sequential data means data that is ordered with respect to some constraint index. For example, time series constitute a popular class of sequential data, where records are indexed by time. It is clear that in temporal data mining it is the ordering among the records is very important and that ordering is the core to the data description/modelling rather than notion of time [3]. Discovery of casual relationships and the discovery of similar patterns within the same time of sequences or among different temporally-oriented events (often called as time series analysis or trend analysis), are the two primarily task of temporal data mining [5]. The supreme goal of temporal data mining is to get wind of hidden relations between sequences and subsequence of events.

One main difference between temporal and conventional time series data mining lies in the size and nature of data sets and the manner in which the data is collected [18]. The second major difference lies in the type of query that we want to estimate or discover from the data [3].

Temporal Data Mining Task:

The possible objectives (or more often we called as 'tasks') of temporal data mining can be classified as Association, Prediction, Classification, Clustering, Characterisation, Search and retrieval, Pattern discovery, Trend analysis and lastly the Sequence Analysis [1].

Classes of Temporal Data:

A. Static Data

Data are called static if all their feature values do not change with time, or change negligibly [16].

B. Sequences

Sequences are commonly referred as ordered sequence of the events or transaction. Though there may not be any explicit reference to time, yet there exists a sort of qualitative temporal relationship (like before, after, during, meet and overlap etc.) between data items.

C. Time Stamped

This category of the temporal data has explicit time related information. Relationship can be quantitative i.e. we can find the exact temporal distance between data element. The consequences obtained through this type of data may be temporal or non temporal in nature.

D. Time Series

Time series data is special case of the time stamped data. In time series data events have uniform distance on the time scale.

E. Fully Temporal

Data of this category is fully time dependent. The inferences are also strictly temporal [1].

III. Literature Survey

Clustering has a long history, with lineage dating back to Aristotle [6]. In our text, we presented some important survey papers on clustering techniques

1. Pedro Pereiva Rodrigous et al. [22] developed an incremental system for clustering streaming time series, using Online Divisive Agglomerative Clustering ODAC system using top-down strategy i.e. hierarchy of clusters. The system uses correlation as similarity measure. It do not need a predefined number of target clusters. It provides a good performance on finding the correct number of clusters obtained by a bunch of runs of k-Means. The disadvantage of this system is when the tree structure expands, the variables should move from root to leaf, when there is no statistical confidence on the decision of assignment may split variables.

2. S. Mishra et al. [17] presented a comparative study based on K-means clustering and agglomerative hierarchical clustering for developing a predictive model for the discharge process. The analysis is carried out in hydrological daily discharge time series of Panchratna station in the river Brahmaputra and Barak Basin Organization in India. The author used Dynamic Time warping (DTW) for measuring similarities in the data.

3. Ramoni et al. [23] The author presented a study on BCD, a Bayesian algorithm for clustering by dynamics. BCD transforms a set S of n numbers of univariate discrete-valued time series into a Markov chain (MC) and then clusters similar MCs to discover the most probable set of generating processes. BCD is basically an unsupervised algorithms based on agglomerative clustering method. The clustering result is evaluated mainly by a measure of the loss of data information induced by clustering, which is specific to the proposed clustering method. They also presented a Bayesian clustering algorithm for multivariate time series [24]. The algorithm searches for the most probable set of clusters given the data using a similarity-based heuristic search method. The measure of similarity is an average of the Kullback–Liebler distances between comparable transition probability tables.

4. Van Wijk and Van Selow [25] in [1999] analyse an agglomerative hierarchical clustering of daily power consumption data based on the root mean square distance. How the clusters distributed over the week and over the year were also explored with calendar-based visualization.

5. Kumar et al. [26] in [2002] presented a distance function based on the assumed independent Gaussian models of data errors and used a hierarchical clustering method to group seasonality sequences into a desirable number of clusters. The experimental results based on simulated data and retail data showed that the new method outperformed both k-means and Ward's method that do not consider data errors in terms of (arithmetic) average estimation error.

6. Vlachos et al. [27] in [2003] introducing a novel anytime version of k-Means clustering algorithm for time series. It is an approach to perform incremental clustering of time-series at various resolutions using the Haar wavelet transform. Using *k-Means* clustering algorithm, for the next level of resolution, they modified the final centers at the end of each resolution as the initial centers. By applying this approach the problem associated with the choices of initial centers for k-Means is completely resolved and it significantly improves the execution time and clustering quality.

7. Li and Biswas [28] the authors described a clustering methodology for temporal data using the hidden Markov model representation. The temporal data are assumed to have Markov property, and may be viewed as the result of a probabilistic walk along a fixed set of (not directly observable) states. The proposed continuous HMM clustering method can be summarized in terms of four levels of nested searches. The HMM refinement procedure for the third-level search starts with an initial model configuration and

incrementally grows or shrinks the model through HMM state splitting and merging operations. They generated an artificial data set from three random generative models: one with three states, one with four states, and one with five states, and showed that their method could reconstruct the HMM with the correct model size and near perfect model parameter values.

8. Bicego, M. Et al. [29] in 2003 studied a novel scheme for HMM based sequential data clustering is proposed, inspired on the similarity based paradigm recently introduced in the supervised learning context. With this approach, a new representation space is built, in which each object is described by the vector of its similarities with respect to a pre-determinate set of other objects. These similarities are determined using hidden Markov models. Clustering is then performed in such a space. By way of this, the difficult problem of clustering of sequences is thus transposed to a more manageable format, the clustering of points (vectors of features). Experimental evaluation on synthetic and real data shows that the proposed approach largely outperforms standard HMM clustering schemes. The main drawback of this approach is the high dimensionality of the resulting feature space, which is equal to the cardinality of the data set.

9. Paredes and Vargas [30] in [2012] their paper presents a novel method to perform clustering of time-series and static data. The method, named Circle-Clustering (CirCle), could be classified as a partition method that uses criteria from SVM and hierarchical methods to perform a better clustering. Different heuristic clustering techniques were tested against the CirCle method by using data sets from UCI Machine Learning Repository. In all tests, CirCle obtained good results and outperformed most of clustering techniques considered in this work. Results showed that CirCle can be used with both static and time-series data.

10. Xiang Lian et al. [31] in [2008] proposed that in all types of time series data, to predict the unknown values that have not arrived at the system and similarity queries based on the predicted data using the three approaches namely Polynomial, discrete Fourier Transform (DFT) and Probabilistic can lead to good offline prediction accuracy but not suitable for online stream environment. Because online requires low prediction and training costs. These approaches are straight forward for seeking general solutions. And it gives proper confidence for prediction. It can predict values while explicitly providing a confidence.

11. Wang et al. [33] Characteristics based clustering of time series data was described by Wang et al. Their paper proposed a method for clustering of time series based on their structural characteristics. Unlike other alternatives, their proposed method does not cluster point values using a distance metric, rather it clusters based on global features extracted from the time series. The feature measures are obtained from each individual series and can be fed into random clustering algorithms, including an unsupervised neural network algorithm, self-organizing map, or hierarchal clustering algorithm. Global measures describing the time series are obtained by applying statistical operations that best capture the underlying uniqueness: trend, seasonality, periodicity, serial correlation, skewness, kurtosis, chaos, nonlinearity, and self-similarity. The empirical results show that their approach is able to yield meaningful clusters.

12. Gong and Richman [32] in [1995]. In their paper, Cluster technique was carried out on a well studied datasets (7-days precipitation data from 1949 to 1987 in central and eastern North America). The Cluster method which they tested were, single linkage, complete linkage, average linkage within a new group, ward's method, k-mean, the nucleated agglomerative method, and the rotated principal component analysis. Similarity measures which they perform are based on three different dissimilarity measure viz. Euclidean, Inverse correlation and theta angle and three initial partitional methods were also tested on the hierarchical and non-hierarchical methods respectively, 22 of 23 cluster algorithms yielded natural grouping solutions.

Results showed that:

- Non- hierarchical methods out performed hierarchical methods.
- The rotated principal component methods were found to be most accurate method.
- The nucleated agglomerated hierarchical method was found to be superior to all other hard cluster methods.
- Ward's method best among hierarchical methods.
- Single linkage always give "chaining" solution, therefore it give poor matching to input data.
- Euclidean Measure, generate more accurate solution.

13. Vernieuwea et al.: The author developed a hydrological modelling of unsaturated groundwater flow based on different data-driven clustering algorithms which are used to identify Takagi–Sugeno models. Takagi-Sugeno models are based on the minimization of an objective function. The Takagi–Sugeno models are identified on the basis of an artificially generated training data set for a specific soil type, and can be incorporated into a fuzzy rule-based groundwater model. They also developed ClusterFinder for guiding the objective function-based clustering algorithms [36].

IV. Clustering Techniques in Data Mining

In this section, we presented a lucid description about various methods of clustering data mining techniques.

Classification of Clustering Data Mining Algorithms:

Clustering methods developed for analysing various static data are classified into five major categories: partitioning methods, hierarchical methods, density based methods, grid-based methods, and model-based methods [16]

- Hierarchical Methods:
- Partitioning Methods
- Density-Based Partitioning Methods:
- Grid-Based Methods
- Model-based methods [16]

It should be remember that for the specificities of time series data, three of above mention clustering (hierarchical, partitioning, and model based) have been applied [19]. But for the readers to enhanced the quality and essentialness of the clustering context, all the five techniques of clustering are discussed in the subsequent subsections.

Hierarchical Methods:

Hierarchical clustering is a method of cluster analysis based on connectivity approach, which seeks to build a hierarchy of cluster. The main idea behind this method is that, element being more related to nearby elements than to elements farther away [8]. Single-link, complete-link, and minimum-variance algorithms are the three major variant of hierarchical

International Journal of Scientific Engineering and Technology (ISSN : 2277-1581)
Volume No.3 Issue No.4, pp : 359-363 1April 2014

clustering, of these three, the single-link and complete-link algorithms are most popular and fashionable [33].In a dendrogram, the elements are represented along the x-axis such that the clusters don't mix, while the y-axis marks the distance at which the clusters merge [8].
The operation of a hierarchical clustering algorithm is illustrated using the two-dimensional data set in Figure 2. The figure shows that seven patterns labelled as P, Q, R, S, T, U, and V, forming three three clusters [12].

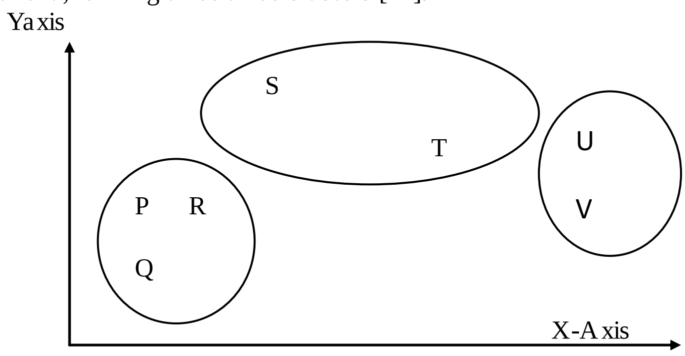

Figure 2. Points falling in three clusters [12].

Agglomerative and divisive are the two most commonly and efficient hierarchical clustering methods [16]. In agglomerative methods, each element is placed in its own cluster and then merging of smaller cluster into bigger cluster is started, until all elements are in a single cluster or until certain termination conditions such as the desired number of clusters are satisfied [16]. Divisive methods do just the opposite of that of agglomerative.
Some of the hierarchical clustering algorithms are: Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), Clustering using representatives (CURE) and CHAMELEON, ROCK [12]. Hierarchical clustering is not restricted to cluster time series with equal length. It is applicable to series of unequal length as well if an appropriate distance measure such as dynamic time warping is used to compute the distance/similarity [16].

Partitioning Algorithms:
Computationally, it is not feasible to check all the possible subsets of the system. So need an approach which will be based on iterative optimization. Partitioning algorithms tries to explore the subset in the dataset at a time. They may also be used in a top down approach. The main function of partitioning algorithms is to divide data into several subsets [4]. The methods in this approach are K-Means, Farthest First Traversal k-center (FFT) algorithm, K-Medoids (PAM), CLARA, CLARANS, Fuzzy K-Means, K-Modes, Fuzzy K Modes, squeezer, K-prototypes, OOLCAT, etc. The most common partitioning method is the K-mean. Some of partitioning algorithms are explained below.
K-Means Algorithms:
It is introduced by J.B.Mac Queen in 1967 and is one of the simplest unsupervised learning algorithms that provide the solution of the well known clustering problem [18]. In this method, entire data set into k subsets such that all points in a given subset are closest to the same centre for an attribute. The K-mean process iterated until there is no change in the gravity centres. The objective function used in measuring the distance between various samples gives the effectiveness of this method [2]. Large data sets is efficient processing by this algorithm, the clusters obtained by this methods have spherical shapes and are likely sensitive to noise [3].
Fuzzy C-Mean:
Fuzzy clustering extends this notion to associate each pattern with every cluster using a membership function; here a data point may belong to more than one cluster producing non disjoint clusters. One widely used algorithm is the Fuzzy C-Means (FCM), which is based on k-means [13]. In this method the affinity of a site to undergo either two or more clusters are visualized. Earlier developed by Dunn and improved by Bezdeck is basically used for pattern recognition.
The objective function

$$Jm = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^{m} \left\|x_i - c_j\right\|^2 , \ 1 \le m \le \infty$$

m= any real number,
u_{ij} = degree of membership of xi in cluster j,
x_i = i^{th} of d-dimensional measured data,
c_j = d-dimension center of cluster [20].

Density Based Clustering: (for Spatial oriented datasets)
Density-based clustering algorithms attempts to find clusters based on density of data points in a given space. The key idea of density-based clustering is that for each instance of a cluster the neighbourhood of a given radius (ϵ) has to contain at least a minimum number of instances (MinPts) [13]. The following points are enumerated as the features of this algorithm.
1. Handles clusters of arbitrary shape
2. Handle noise
3. Needs only one scan of the input dataset.
4. Needs density parameters to be initialized [7].

There are two major approaches for density-based methods. The first approach pins density to a training data point and is reviewed in the subsection Density-Based Connectivity. Representative algorithms include DBSCAN, GDBSCAN, OPTICS, and DBCLASD. The second approach pins density to a point in the attribute space and is explained in the subsection Density Functions. It is represented by the algorithm DENCLUE that is lesser affected by data dimensionality [10] [13].

Grid Based Algorithms:
Grid-based algorithms generally have a fast processing time as compared with the existing clustering algorithms. This algorithm first operates a uniform grid to collect the regional statistic data and, then, perform the clustering on the grid, instead of the database directly [9]. Grid based methods subdivide the object space into a finite number of cells (hyper-rectangles) and then perform the required operations on the quantized space [15]. The performance of grid-based approach normally depends on the size of the grid which is usually much less than the database. However, for highly irregular data distributions, using a single uniform grid may not be sufficient to obtain a required clustering quality or fulfill the time requirement [9]. The representative grid-based clustering algorithms are STING, WaveCluster, CLIQUE and MAFIA [4].

Model-based methods:

Model-based methods assume a model for each of the clusters and attempt to best fit the data to the assumed model. There are two major approaches of model-based methods: statistical approach and neural network approach. An example of statistical approach is AutoClass, which uses Bayesian statistical analysis to estimate the number of clusters. Two prominent methods of the neural network approach to clustering are competitive learning, including ART and self-organizing feature maps [16].

V. Conclusion

In last two decades, information gathering at a very faster rate and analysing these information to find some interestingness is of key interest. Data mining is that powerful technology which extracts the hidden interestingness from such databases. In which, cluster analysis is utmost importance. The ubiquitous nature of temporal datasets led to an extension of the scope of new trends and methods of data mining techniques. Lot of research work has been carried on this field to develop more mature and efficient clustering algorithms for temporal data mining. In this paper, we surveyed the current studies on temporal data sets clustering. There are numerous methods and algorithms developed for clustering techniques, which are discussed in very unique and lucid style. Along with this, applications of various cluster algorithms in many fields ranging from economy, medical surveillance to stock market, in which applications on hydrology is profoundly mentioned.

Also, a brief discussion of temporal data mining along it various tasks and phenomena makes this paper a vital literature. Since the clustering analysis itself a vast area for scientific research and by each day yet to set, either a new algorithms are discovered or developed with some improvement over the existing one. We thus outlined monotonically only the key concepts and applications for cluster techniques.

References

I. *Mohd. Shahnawaz et al.: Temporal Data Mining: An Overview, IJEAT, Volume-1, Issue-1, October 2011.*

II. *Kavitha, V. And.Punithavalli, M.: Clustering Time Series Data Stream – A Literature Survey, IJCSIS, Vol. 8, No. 1, April 2010.*

III. *Laxman, S. And Sastry, P.S.: A survey of temporal data mining, Sadhana, Vol. 31, pp. 173-198, 2006.*

IV. *Mann, A. K. et al.: Survey paper on clustering techniques, IJSETR, Vol. 2, Issue 4, April 2013.*

V. *Geetha, R. et al: A survey of spatial, temporal and spatio-temporal data mining, Journal of Computer Application, Vol. 1, 2008.*

VI. *Xu, R. and Wunsch, D. II : Survey of Clustering Algorithms, IEEE transactions on neural networks, vol. 16, no. 3, May 2005.*

VII. *M.Vijayalakshmi, M.Renuka Devi: A Survey of Different Issue of Different clustering Algorithms Used in Large Data sets, IJARCSSE, Volume 2, Issue 3, March 2012.*

VIII. *Jain, A. K., and Maheshwari, Satyam: Survey of Recent Clustering Techniques in Data Mining, IJCSMR, Vol 1, Issue 1,p.p. no . 72-78, Aug 2012.*

IX. *Liao, W. K. et. al. : A Grid-based Clustering Algorithm using Adaptive Mesh Refinement, Appears in the 7th workshop on Mining scientific and Engineering Datasets, 2004.*

X. *Berkhin, P. : A survey of clustering data mining techniques, pp 25-71, 2002.*

XI. *Gersho, A. and Gray, R. M.: Vector quantization and signal compression, Kluwer Academic Publishers, 1992.*

XII. *JAIN, A.K. : Data Clustering: A Review, ACM Computing Surveys, Vol. 31, No. 3, September 1999.*

XIII. *Rama. B et al: A Survey on clustering Current status and challenging issues, IJCSE, Vol. 02, No. 09, 2010.*

XIV. *Sharma, Ritu et al.: K-Means Clustering in Spatial Data Mining using Weka Interface, (ICACACT) 2012 Proceedings published by IJCA.*

XV. *Ilango, M. R. and Mohan, V.: A survey of Grid based clustering algorithms, IJEST, Vol. 2, 8, 2010, p.p. 3441-3446.*

XVI. *Liao, T. W.: Clustering of time series data—a survey, Pattern Recognition Society, 2005.*

XVII. *Mishra, S. Et al. : Pattern Discovery in Hydrological time series data mining during the monsoon period of the high flood years in Brahmaputra River basin, IJCA, vol. 67, no. 6., 2013.*

XVIII. *Rani ,Y. L. S. et al.: Clustering algorithms for temporal data mining: An overview, IJETAE, Volume 3, Issue 7, July 2013.*

XIX. *Esling, P. And Agon, C.: Time series data mining, ACM Computing Surveys, Vol V.*

XX. *Kar, A. K. et al: Development of Flood Forecasting systems using statistical and ANN techniques in the downstream catchment of Mahanadi Basin, India, J. Water Resource and Protection, 2010, p.p. 880-887.*

XXI. *Mcgarry, Ken: A Survey of interestingness measures for knowledge discovery, The Knowledge Engineering Review, p.p. no. 1-24, 2005.*

XXII. *Rodrigues, P. et al.: "Hierarchical Clustering of Time Series Data Streams," IEEE Transactions on Knowledge and Data Engineering, vol. 20, no. 5, pp. 615-627, May 2008.*

XXIII. *Ramoni, M. et al.: Bayesian clustering by dynamics, Mach. Learning 47 (1) (2002) 91–121.*

XXIV. *Ramoni, M. et al.: Multivariate clustering by dynamics, Proceedings of the 2000 National Conference on Artificial Intelligence (AAAI-2000), San Francisco, CA, 2000, pp. 633–638.*

XXV. *Wijk, J. J. van and Selow, E. R. Van: Cluster and calendar based visualization of time series data, Proceedings of IEEE Symposium on Information Visualization, San Francisco, CA, October 25–26, 1999.*

XXVI. *Kumar, M. et al. :Clustering seasonality patterns in the presence of errors, Proceedings of KDD '02, Edmonton, Alberta, Canada.*

XXVII. *Vlachos, M. et al.: A wavelet based anytime algorithm for k-means clustering of time series, Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, CA, 2003.*

XXVIII. *Li, C. And Biswas, G.: Temporal pattern generation using hidden Markov model based unsupervised classification, Lecture Notes in Computer Science, vol. 164, IDA '99, Springer, Berlin, 1999, pp. 245–256.*

XXIX. *Bicego, M. et al.: Similarity Based Clustering of Sequences Using Hidden Markov Models, Springer-Verlag Berlin Heidelberg, pp. 86–95, 2003.*

XXX. *Paredes, G. E. and Vargas, L. S., : Circle-Clustering: A New Heuristic Partitioning Method for the Clustering Problem, WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012.*

XXXI. *Lian, X. et al.: Efficient similarity search over future stream time series, IEEE transaction on Knowledge and Data Engineering, vol. 20, no. 1, p.p. 40-54, 2008.*

XXXII. *Gong, X and Richman, M. B. ,: On the application of cluster analysis to growing season precipitation data in North America East of the Rockies, Journal Of Climate, Vol. 8, April 1995, p.p. no. 897-931.*

XXXIII. *Wang, X. et al.: Characteristic-Based Clustering for Time Series Data, Journal on Data Mining and Knowledge Discovery, vol. 13, no. 3, pp. 335-364, 2006.*

XXXIV. *Pratihar, D.K. and Das, A.: Soft Computing, Narosa Publishing House, 2009.*

XXXV. *Wang, C. W. et al. : A comparison of performance of several artificial intelligent methods for forecasting monthly discharge time series, Journal of Hydro., Vol. 374, No. 3-4, pp. 294-306, 2009,.*

XXXVI. *Vernieuwea, H.:Comparison of clustering algorithms in the identification of Takagi–Sugeno models: A hydrological case study, Fuzzy Sets and Systems, vol.157 , p.p. no. 2876 – 2896, 2006.*