Time Series Analysis : A Review of methods

Diana Moses
Department of Computer Science and Engineering,
St. Peter's Engineering College
Hyderabad, India

***Abstract*— Time series is an essential class of transient information articles and it very well may be effectively gotten from logical and budgetary applications. A time series is an accumulation of perceptions made sequentially. The idea of time series information incorporates: substantial in information estimate, high dimensionality and important to refresh constantly. In addition time series information, which is described by its numerical and nonstop nature, is constantly considered all in all rather than individual numerical field. The expanding utilization of time series information has started a lot of innovative work endeavors in the field of information mining. The copious research on time series information mining in the most recent decade could hamper the section of intrigued analysts, because of its unpredictability. In this paper, an extensive modification on the current time series information mining research is given. They are commonly classified into portrayal and ordering, comparability measure, division, perception and mining. Besides best in class examine issues are additionally featured. The essential goal of this paper is to fill in as a glossary for intrigued scientists to have a general picture on the ebb and flow time series information mining advancement and distinguish their potential research course to advance examination.**

Keywords—Time Series Analysis, Classification, Segmentation, Pattern Matching, Subsequence matching

I. INTRODUCTION

As of late, the expanding utilization of transient information, specifically time series information, has started different innovative work endeavors in the field of information mining. Time series is an imperative class of transient information articles, and it tends to be effectively acquired from logical and monetary applications (for example electrocardiogram (ECG), day by day temperature, week after week deals sums, and costs of common assets and stocks). A time series is a gathering of perceptions made sequentially. The idea of time series information incorporates: extensive in information measure, high dimensionality and refresh ceaselessly. Also time series information, which is described by its numerical and consistent nature, is constantly considered all in all rather than individual numerical field.

Thus, dissimilar to conventional databases where similarity look is definite match based, similarity seek in time series information is regularly completed in a rough way. There are different sorts of time series information related research, for instance, finding comparative time series (1-3), subsequence looking in time series (4), dimensionality decrease (5-8) and division (9). Those inquires about have been concentrated in significant detail by both database and example acknowledgment networks for various spaces of time series information (5-8). With regards to time series information mining, the crucial issue is the way to speak to the time series information. One of the basic methodologies is changing the time series to another space for dimensionality decrease pursued by an ordering system.

Furthermore, similarity measure between time series or time series subsequences and division are two center undertakings for different time series mining errands. In light of the time series portrayal, diverse mining errands can be found in the writing and they can be generally ordered into four fields: design disclosure and bunching, characterization, rule revelation and rundown. A portion of the paper focuses on one of these fields, while the others may concentrate on more than one of the above procedures. In this paper, an exhaustive survey on the current time series information mining research is given. Three cutting edge time series information mining issues, gushing, multi-trait time series information and protection are additionally quickly presented.

II. LITERATURE SURVEY

One of the real explanations behind time series portrayal is to lessen the measurement (for example the quantity of information point) of the first information. The most straightforward technique maybe is sampling (10). In this strategy, a rate of l/n is utilized, where l is the length of a time series T and n is the measurement after dimensionality reduction (Fig. 1).

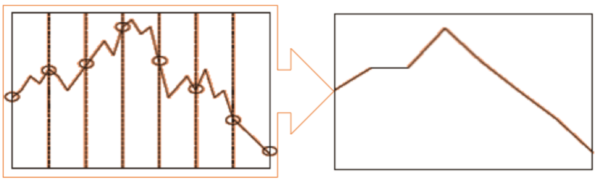

Figure 1. Sampling and distortion

Notwithstanding, the sampling strategy has the downside of mutilating the state of tested/packed time series, if the sampling rate is excessively low. unique information. An improved technique is to utilize the average(mean) estimation of each fragment to speak to the comparing set of information focuses. Once more, with time series $T = (T_1, T_2, \dots .T_l)$ and n is the measurement after dimensionality reduction, the "compacted" timeseries can be acquired by

$$\widehat{p_k} = \frac{1}{e_k - s_k + 1} \sum_{i=s_k}^{e_k} p_i$$

where s_k and e_k indicate the beginning and completion information points of the k^{th} portion in the timeseries P, separately. That is, utilizing the sectioned way to speak to the timeseries (11). This technique is likewise called piecewise aggregate approximation (PAA) by Keogh et al. (5,6).

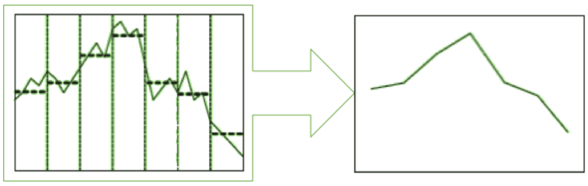

Figure 2. Piecewise Aggregate Approximation

Keogh et al. proposed an all-inclusive adaptation called an adaptive piecewise constant approximation (APCA), in which the length of each section isn't settled, yet adaptive to the state of the series. A mark system is proposed by Faloutsos et al. with comparable thoughts. Other than utilizing the intend to speak to each portion, different techniques are proposed (4,12,13). For instance, Lee et al. proposed to utilize the sectioned total of variety (SSV) to speak to each fragment of the timeseries (14,15). Moreover, somewhat level approximation is proposed by Ratanamahatana et al. and Bagnall et al, which utilizes a bit to speak to every datum point (16-21). To lessen the element of timeseries information, another methodology is to inexact a timeseries with straight lines. Two noteworthy classifications are included. The first is linear interpolation. A typical technique is utilizing piecewise linear representation (PLR) (7,8). The approximating line for the subsequence $P(p_i, p_j)$ is basically the line associating the information points p_i and p_j. It tends to intently adjust the end purpose of back to back fragments, giving the piecewise approximation with associated lines. PLR is a base up calculation. It starts with making a fine approximation of the time series, so l/2 sections are utilized to rough the l length time series and iteratively combines the most minimal cost pair of fragments, until it meets the required number of portion. At the point when the pair of contiguous fragments S_i and S_{i+1} are blended, the expense of combining the new portion with its correct neighbor and the expense of consolidating the S_{i+1} section with its new bigger neighbor is determined and stretches out PLR to various leveled structure. Besides, Keogh and Pazzani upgrade PLR by thinking about loads of the sections and importance criticism from the user (8). The second methodology is linear relapse, which speaks to the subsequences with the best fitting lines (22,23). Besides, lessening the measurement by safeguarding the notable points is a promising strategy. These points are called as perceptually important points (PIP) as shown in Fig 3. The PIP distinguishing proof procedure is first presented by Chung et al (2001) and utilized for example coordinating of specialized (examination) designs in monetary applications (31). With the timeseries T, there are n information points: T1, T2, Tn. Every one of the information points in P can be reordered by its significance by experiencing the PIP distinguishing proof procedure. The main information point T1 and the last information point Tn in the timeseries are the first and two PIPs, separately. The following PIP that is discovered will be the point in P with most extreme separation to the initial two PIPs. The fourth PIP that is discovered will be the point in T with most extreme vertical separation to the line joining its two contiguous PIPs, either in the middle of the first and second PIPs or in the middle of the second and the last PIPs.

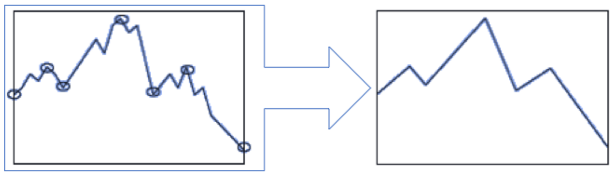

Figure 3. Perceptually Important Points (PIP).

The PIP area process proceeds until every one of the points in P are joined to are requested rundown L or the required number of PIPs is come to (for example diminished to the required measurement). Seven PIPs are distinguished in from the example timeseries in Fig. 3. Point by point treatment can be found (24-26). The thought is like a procedure proposed around 30 years back for lessening the quantity of points required to speak to a line by Perng et al (25) utilize a milestone model to distinguish the important points in the timeseries for similarity measure. Then proposed a cross section structure to speak to the distinguished pinnacles and troughs (called control points) in the timeseries and characterize extrema as minima and maxima in a timeseries and pack the time by choosing just certain important extrema and dropping alternate points. The thought is to dispose of minor variances and keep major minima and maxima.

III. SIMILARITY MEASURE

Similarity measure is of basic significance for an assortment of time series examination and information mining undertakings. The greater part of the representation approaches talked about in Section 2 additionally propose the similarity measure technique on the changed representation plot. In customary databases, similarity measure is precise match based. Anyway in time series information, which is described by its numerical and constant nature, similarity measure is regularly done in an in exact way.

In timeseries area, conceiving a fitting similarity work is in no way, shape or form unimportant. There are basically two different ways the information that may be sorted out and handled (1). In whole sequence matching, the whole length ever series is considered amid the similarity look. It requires contrasting the question sequence with every hopeful series by assessing the separation capacity and monitoring the sequence with the littlest separation. In subsequence matching, where a question sequence Q and a more extended sequence P are given, the assignment is to discover the subsequences in P, which matches Q. Subsequence matching necessitates that the question sequence Q be set at each conceivable counterbalance inside the more drawn out sequence P. Megalooikonomou et al think about the helpfulness of various similarity measures for bunching comparative stock timeseries (26,27).

A. *Whole sequence matching*

To measure the similarity/dissimilarity between two time series, the most famous methodology is to assess the Euclidean separation on the changed representation like the DFT coefficients (1) and the DWT coefficients Chan and Fu.

Albeit the greater part of these methodologies ensure that a lower bound of the Euclidean separation to the first information, Euclidean separation isn't continually being the appropriate separation work in indicated spaces (5, 25-27). For instance, stock time series has its own qualities over other time series information (for example information from logical territories like ECG), in which the striking points are important.

Without a doubt, a standout amongst the most well known and field-tried similarity measures is known as the "time warping" remove measure. In light of the dynamic time warping (DTW) strategy, the proposed technique in (2) predefines a few examples to fill in as layouts with the end goal of example recognition. To adjust two timeseries, P and Q, utilizing DTW, a n-by-m framework M is first built. The (i_{th}, j_{th}) component of the lattice, m_{ij}, contains the separation $d(q_i, p_j)$ between the two points q_i and p_j and an Euclidean separation is regularly utilized, for example $d(q_i,p_j).(q_i\_p_j)$. It relates to the arrangement between the points q_i and p_j. A warping way, W, is a touching arrangement of framework components that characterizes a mapping among Q and P. Its kth component is characterized as $w_k.(i_k, j_k)$ and
$W = w_1, w_2, \ldots, w_k, \ldots, w_K$.
where $\max(m, n) \leq K \leq m+n-1$

The warping way is normally exposed to the accompanying imperatives. They are boundary conditions, continuity and monotonicity. Boundary conditions are $w_1= (1,1)$ and $w_k= (m,n)$. This requires the warping way to begin and complete corner to corner. Next imperative is continuity. Given wk, at that point $w_k=(a,b)$, where $a-a' \leq 1$ and $b-b' \leq 1$. This limits the permissible strides in the warping way being the neighboring cells, including the corner to corner contiguous cell. Likewise, the imperatives $a-a' \geq 0$ and $b-b' \geq 0$ compel the points in W to be monotonically dispersed in time. There is an exponential number of warping ways fulfilling the above conditions. In any case, just the way that limits the warping cost is of intrigue. This way can be effectively found by utilizing dynamic programming to assess the accompanying repeat condition that characterizes the total separation g as the separation d found in the present cell and the base of the combined separations of the neighboring components.

B. Subsequence matching

In subsequence matching where a question sequence and a more extended time series are given, the undertaking is to discover the subsequences in the more drawn out time series, which coordinates the inquiry sequence. The question sequence is required to put at each balance inside the more drawn out time series. Faloutsos et al. sums up the work in Agrawal et al. for subsequence seeking. In view of this work, numerous inquires about are directed to enhance the execution of the subsequence looking (1,4). For instance, a DualMatch is proposed by Moon et al. (2001) to partition the time series into disjoint windows and inquiry design into sliding windows. Loh and Kim (2001) expanded Faloutsos et al., utilizing a list interpolation to explain the capacity and CPU time overhead (4). The GeneralMatch technique is proposed by Moon et al. (2002), which decreases the window estimate impact by utilizing huge windows by the technique in Faloutsos et al. and misuses point-sifting impact by DualMatch (4). Besides, Kim and Jeong (2007) talk about on the potential execution bottleneck amid subsequence matching. Four noteworthy territories, that time required to process subsequence matching, are distinguished. They are handling time, plate get to time and the comparing post-preparing ventures of them. A window requesting strategy is proposed to dispense with the redundancies of circle access and CPU preparing in the post-handling steps. A strategy dependent on a list interpolation is additionally proposed by Lim et al. (2007) to enhance the execution of DualMatch.

IV. SEGMENTATION

Time series segmentation can be considered either as a preprocessing venture for assortment of information mining assignments or as trend analysis procedures. It is additionally considered as a discretization issue. Not at all like value-based databases with discrete things, time series information is described by their numerical and constant nature. In Das et al. (1998), a basic discretization technique is proposed. A settled length window is utilized to portion a time series into subsequences and the time series is then spoken to by the crude shape patterns that are framed. This discretization process basically relies upon the decision of the window width. Be that as it may, utilizing settled length segmentation is an over-disentangled way to deal with tackle the issue. There are something like two recognized impediments. To start with, important patterns ordinarily show up with various lengths all through a time series. Second, because of the even segmentation of a time series, significant patterns might be missed on the off chance that they are part crosswise over time (cutting) points. Accordingly, it is smarter to utilize a dynamic methodology, which distinguishes the time points in an increasingly adaptable manner (for example utilizing diverse window widths). This is absolutely not a paltry segmentation issue. Basic segmentation techniques incorporate utilizing the PIP (Fu et al., 2006; Jiang et al., 2007) or identifying unique occasions (Guralnik and Srivastava, 1999) in the time series as the time points, least message length (MML) (Oliver et al., 1998) and least depiction length (MDL) segmentation (Fitzgibbon et al., 2002). Fancourt and Principe (1997) receive PCA for the segmentation issue. In light of PCA, a fluffy bunching based segmentation is proposed by Abonyi et al., (2003, 2005). A two phases approach which first uses piecewise generalized likelihood ratio (GLR) to unpleasant segmenta-tion and after that refines the outcomes is proposed by Wang and Willett (2004). Then again, Keogh et al. (2001c) embrace PLR to fragment the time series. They center around the issue of an online segmentation of time series and a sliding window and base up (SWAB) approach is proposed.

The segmentation issue has likewise been considered from the point of view of finding cyclic periodicity for the majority of the portions. In Han et al. (1998, 1999), the information 3D square and the Apriori information mining procedures are utilized to mine section savvy periodicity, utilizing a settled length period. A disconnected procedure for the aggressive distinguishing proof of piecewise stationary time series is depicted by Fancourt and Principe (1996). Notwithstanding performing piecewise segmentation and recognizable proof, the proposed techni-que maps comparable sections of a time series as neighbors on an area map. Himberg et al. (2001) propose a worldwide iterative substitution (GIR) technique, which approximates the dynamic programming result for limiting the intra section fluctuations. The proposed technique is connected to setting acknowledgment for the cell phone applications. In spite of the fact that the methodologies portrayed in this area can commonly distinguish a given pattern from a time series, they don't think about the issue of recognizing an appropriate arrangement of time points in a time series, when a lot of pattern formats is given; for instance, the specialized patterns (for example H&S, twofold best, and so on.) for the stock analysis. Further, so as to frame an adaptable mining space, an assortment of patterns (for example in various goals) must be distinguished. The previously mentioned segmentation errand can be viewed as an improvement issue and Chung et al. propose an answer, which depends on a developmental calculation (31,32).

V. VISUALIZATION

Visualization is an important system to introduce the processed time series for further analysis by clients. It is likewise an integral asset to encourage the mining errands like pattern seeking, question by-precedent, and pattern revelation subsequently. Current apparatuses for envisioning time series include: (1) bunch and date-book based visualization device of van Wijk and van Selow, 1999, which acquires pieces of information with a given interim and after that groups them in like manner and winding visualization device of Weber et al (33-35), which maps the intermittent area of time series into a ring. These two apparatuses are centered around intermittent time series and a settled length of period must be given, state, week by week or month to month. A budgetary visual investigation framework for pattern-based analysis of 2-dimensional time-shift graph information is proposed. Hao et al present the thought of level of intrigue (DOI) to characterize and generate multi-goals formats of long time series. Non-linear rescaling and space-effective rendering strategy are utilized to picture the long time series (36).

As of late, another timeseries visualization device called VizTree is proposed (37-41). This methodology first proselytes each numeric timeseries to a symbol string dependent on the SAX and a lot of substrings (with a similar number of symbol) extricated from the symbol string is encoded by an adjusted addition tree to picture the recurrence of patterns. That is, the SAX discretizes the first timeseries into settled length subsequences, changes over every subsequence to a symbol and the symbols got are connected to shape a symbol string. Given a symbol string, say abccbccbcabcbcc, the following stage is to change over this long string to a lot of substrings as per the length of every substring, W (or the quantity of substring),specified by clients.

VI. TIME SERIES DATA MINING

Mining is the last objective to find hidden data or learning from either the first or the changed time series information. To be sure, pattern disclosure is the most well-known mining errand and the grouping technique is the most generally strategy. Other time series information mining assignments incorporate order, rule mining and outline.

It is a non-trifling assignment to find intriguing patterns, which incorporate as often as possible showing up Fu et al (42-46) and amazing patterns (6), from time series information. These errands are additionally called theme disclosure by Chiu et al and Tanaka et al (47,48) and oddity discovery or discovering dissensions Keogh et al (7), separately. The revelation of intriguing patterns has turned out to be a standout amongst the most important information mining assignments, and it tends to be connected to numerous areas of Caraa-Valente and Lopez-Chavarrias, and Lerner et al (49,50)

He also presented a support vector relapse (SVR)- based online oddity recognition calculation. Chan and Mahoney (51) present an online oddity discovery approach dependent on the Gecko calculation, which makes a sequence of insignificant jumping boxes with the training directions.

Autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models have likewise been utilized widely for time series analysis. Kalpakis et al (52) propose to group ARIMA time series, utilizing the segment around the medoids technique. Xiong and Yeung (2004) center around the issue of bunching time series of various lengths, utilizing blends of ARMA models and desire augmentation (EM) calculation. Bagnall and Janacek (20,21) center around bunching information got from ARMA models, utilizing k-means and k-medoids calculations. A cut-out process, which descretizes information into double sequences of above and beneath the media, is received. This process is fortifies on the nearness of an exception in the information. Hidden Markov model (HMM) is a typical model-based calculation embraced in timeseries bunching by Panuccio et al (53). Its characterized as stochastic speculations of limited state automata, where the two advances among states and generation of yield symbols are administered by likelihood conveyances. Oates et al (56) present a crossover timeseries grouping calculation that utilizes DTW for harsh introduction and HMM for removing the sequences that don't have a place with the bunches. Yin and Yang(2005) propose to change the sensor timeseries information into an equivalent length vector and model it as a HMM for otherworldly bunching. Besides, are cursive HMM training process is proposed by Duan et al (57).

A. *Classification*

Classification is a customary data mining undertaking. In the time series area, unique treatment must be considered because of the idea of the data. Geurts (58) proposes to arrange time series data dependent on joining nearby properties or patterns in the time series. Zhang et al (59,60) build up a representation strategy utilizing wavelet deterioration that can naturally choose the parameters for the classification errand. They propose a closest neighbor classification calculation, utilizing the determined proper scale. Kadous and Sammut (61) use metafeature approach (for example repeating substructure) like neighborhood maxima in time series to generate classifiers. Additionally, Yang et al. (62) center around highlight subset choice (FSS) in light of regular principal components, which is called CleVer, to hold the relationship data among unique highlights. Classification is utilized to assess the viability of the chose subset of highlights. Then again, analysts have additionally centered around modifying or creating classifiers for time series data. For instance, Povinelli et al. (63-65) present a flag classification approach dependent on modeling the dynamics of a framework as they are caught in a remade stage utilizing Gaussian Mixture models of time space marks. Rodriguez and Alonso (66) ponder both interim and DTW-based choice trees satisfactory for the classification of time series data. Outfits are utilized to join base classifiers, while others ponder the blend of the numerosity reduction, utilizing DTW and closest neighbor classifiers for time series classification. Likewise, its proposed a semi-directed time series classifiers when just a little arrangement of named models is accessible.

VII. FURTHER RESEARCH DIRECTIONS

Because of the develop advancement in this field and the critical improvement on the equipment and correspondence advances, three expansions pull in more scientists concentrated on as of late. They are mining on multi-characteristic time series, mining on time series data stream and furthermore the security issue. A few explores talked about above additionally proposed halfway arrangements or headings on them. Initially, multi-characteristic time series data can likewise be considered as numerous time series control. Povinelli and Feng (65) propose a methodology which transient groups from various time series are utilized and a hereditary calculation is received. The strategy reproduces state space for transient pattern extraction and embraces an ideal nearby model for momentary determining. The execution of the methodology is demonstrated by utilizing budgetary non-stationary time series (for example stock cost and volume). Kahveci et al. (67,68) think about the issue of move and scale invariant scan for multi-property time series. A symmetric separation work and a Cone Slice (CS) record are proposed with center around mining of shut patterns in multi-sequence time series by receiving a SAX representation.

In light of the all-around created timeseries datamining calculations in various viewpoints, they are either connected straightforwardly or altered for spilling timeseries data. In fact, representation of spilling timeseries for dimensionality reduction and an online question or matching is a hotly debated issue. It is important that an approaching stream of data is a constantly attached timeseries in a database. Each time when another data point arrives, the framework needs to get/get the data from the database, the closest or the neighboring data of the approaching timeseries is upto the time position and most inquires about spotlight on researching the connection of the data. Yi et al (11-13) build up a quick strategy to break down the co-advancing timeseries for evaluating and guaging, quantitative datamining and exception location. Gilbert et al. (69) receive sketch based techniques for catching different linear projections of the data for speaking to datastreams (i.e.wavelet change) and rough aggregate question. Diana et al used timeseries to analyze ECG for detection abnormal patterns in the human ECG (70-72). Gao and Wang (73,74) handle the issue by utilizing a FFT to locate the cross relationships of timeseries in a clump mode proficiently. Gao et al. (73,74) center around nonstop closest neighbor inquiry, utilizing existing ordering techniques with pre-getting. Gao et al. (73,74) propose to join a few basic strategies (e.g.sketches, convolution, organized irregular vectors,etc.) to register Pearson relationship over uncooperative timeseries. Vlachos et al. (75) inspect the issue of observing and distinguishing relationship burst patterns in multi-steam timeseries data. The arrangement embraces burst location and ordering. Yi et al (11-13) built up a change based structure to lessen the measurement for vast scale and dynamic timeseries data on the web. The structure is centered around DFT-based summary generation and a recursive technique is acquainted with refresh the most elevated vitality change coefficients of the series data. Wei et al .(41) present an on-the-fly subsequence matching of gushing time series to a lot of predefined patterns, utilizing the sifting approach. The proposed methodology consolidates comparable patterns into a wedge, which is an envelope-based lower bouncing strategy, to speedup the matching process. A stream-DTW(STW) separate measure is proposed for consistently observing DTW remove measure of timeseries data streams. Boolean representation dependent on the data-adaptive connection analysis proposed by Zhang et al. (59-60). Furthermore, they present subjective client determined amnesic capacities dependent on PLR to permit an online approximation of spilling time series. This capacity permits discretionary, client characterized reduction of value with time. A tree structure is additionally proposed by Fu et al. (42-46) for putting away the PIPs, which bolsters different gradual refreshing approaches of Fu et al. (42-46). A multiscale segment mean (MSM) approximation is proposed which support gradually calculation and static/dynamic pattern matching.

An instructional exercise introducing strategies for finding sliding window relationships, finding blasts, matching murmurs and keeping up and controlling time requested datastream can be found in Lerner et al.(50). Chan et al. (51) propose to deal with ceaseless pattern discovery, utilizing spatial amassing separation (SpADe). A SpADe is proposed in this paper to deal with both moving and scaling in transient and plentifulness measurements. Lin et al.(37-41) concentrate on persistent inquiry sequences on time series datastream dependent on window development component for supporting variable length questions. Two online segmentation techniques, (stepwise) doable space window (FSW/SFSW),are proposed by Lin et al. (37-41) to enhance the execution of great sliding window strategy. A twisting free prescient spilling timeseries matching calculation is presented by Loh et al.(28-30). The proposed calculation performs preprocessing venture to evacuate twists and foresee future list items at the same time.

Besides, analysts likewise expand their enthusiasm on mining time series gushing data. A web based learning structure dependent on a probabilistic model for exception location and change-point recognition on the time series datastream is also proposed. A spilling pattern revelation strategy in various timeseries which outlines the keytrends in the stream accumulation dependent on PCA was presented. An internet bunching framework is proposed by Rodrigues et al.(66) which on down strategy is embraced to develop a parallel tree order of groups. Bunches' distances across are developed persistently with the streamdata.

Third, look into on datamining is proposed to incorporate with protection concern (1). Dealing with the security in timeseries datamining is a recently explore heading. Zhu et al. (50) recommend that customary strategies are not compelling in the timeseries data. Data stream separation assault is recognized and conceivable counter measures to this assault are additionally proposed in this paper. To protect the security, shrouded time series data are recommended to be embraced. Lin et al.(37-41) propose a way to deal with manage shrouded range query(CRQ) in view of a R-tree ordering. Besides, structure is to assess diverse timeseries assurance techniques. A lot of data misfortune and revelation is k measures for time series are presented in this paper. In light of the meaning of these sorts of measurements, expanding number of research on timeseries data assurance and security issue amid the mining process is normal.

VIII. Conclusion

In this paper, we have checked on research in time series data mining. Diverse research is centered around at least one issues in time series data mining. Be that as it may, as indicated by the novel conduct of the time series data, existing examination is as yet deficient and it is considered as one of the 10 testing issues in data mining. There is still space for us to additionally research and create. For instance, while the vast majority of the examination networks have concentrated on the mining errands, the essential issue on the best way to speak to a time series has not yet been completely tended to up until now. To speak to a time series is fundamental, since time series data is difficult to control in its unique structure. The high dimensionality of time series data makes challenges in applying existing data mining techniques to it. In this way, characterizing a progressively powerful and productive time series representation conspire is of basic significance.

REFERENCES

[1] Agrawal, R., Faloutsos,C., Swami,A., 1993. Efficient similarity search in sequence databases. In Proceedings of the Fourth International Conference on Foundations of Data Organization and Algorithms, pp.69–84.

[2] Berndt, D.J., Clifford,J., 1996. Finding patterns in timeseries : a dynamic programming approach. Advances in Knowledge Discovery and Data Mining, 229–248.

[3] Chan, K.P., Fu,A.C., 1999. Efficient timeserie matching by wavelets. In Proceedings of the 15th IEEE International Conference on Data Engineering, pp. 126–133.

[4] Faloutsos, C., Ranganathan,M., ManolopoulosY., 1994. Fast subsequence matching in time-series databases. In Proceedings of the 1994 ACM SIG MOD International Conference on Management of Data, pp.419–429.

[5] Keogh, E., 1997. A fast and robust method for pattern matching in timeseries databases. In Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence, pp.578–584.

[6] Keogh, E., 1997. Fast similarity search in the presence of longitudinal scaling in time series databases. In Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence, pp.578–584.

[7] Keogh, E., 2002.Exact indexing of dynamic time warping. In Proceedings of the 28th International Conference on Very Large Databases, pp.406–417.

[8] Keogh, E., Kasetty,S., 2002.On the need for timeseries data mining benchmarks: a survey and empirical demonstration. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 102–111.

[9] Abonyi, J., Feil,B., Nemeth,S., Arva,P., 2005.Modified Gath–Geva clustering for fuzzing segmentation of multivariate time-series. Fuzzy Sets and Systems, Data Mining Special Issue, 149,39–56.

[10] Astrom, K.J., 1969. On the choice of sampling rates in parametric identification of time series. Information Sciences, 1 (3), 273–278.

[11] Yi, B., Faloutsos,C., 2000. Fast time sequence indexing for arbitrary Lpnorms. In Proceedings of the 26th International Conference on Knowledge Discovery and Data Mining Very Large Data Bases, pp. 385–394.

[12] Yi, B., Jagadish,H.V., Faloutsos,C., 1998. Efficient retrieval of similar time sequences under time warping. In Proceedings of the 14th IEEE International Conference on Data Engineering, pp.201–208.

[13] Yi, B.K., Sidiropoulos,N.D., Johnson,T., Jagadish,H.V., Faloutsos,C., Biliris,A., 2000. Online data mining for co-evolving time sequences. In Proceedings of the 16th IEEE International Conference on Data Engineering, pp.13–22.

[14] Lee, A.J.T., Wu,H.W., Lee,T.Y., Liu,Y.H., Chen,K.T., 2009.Mining closed patterns in multi-sequence time-series databases. Data and Knowledge Engineering, 68 (10), 1071–1090.

[15] Lee, S., Kwon,D., Lee,S., 2003.Dimensionality reduction for indexing timeseries based on the minimum distance. Journal of Information Science and Engineering 19,697–711.

[16] Ratanamahatana, C.A.and Keogh,E., 2004. Making time-series classification more accurate using learned constraints. In Proceedings of the Fourth SIAM International Conference on Data Mining, pp.11–22.

[17] Ratanamahatana, C.A., Keogh,E., 2005. Three myths about dynamic time warping data mining. In Proceedings of the Fifth SIAM International Conference on Data Mining.

[18] Ratanamahatana, C.A., Keogh,E., Bagnall,A.J., Lonardi,S.A,2005.Novel bit level

[19] time series representation with implications for similarity search and clustering. In Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp.771–777.

[20] Bagnall, A., Janacek,G., 2005.Clustering timeseries with clipped data. Machine Learning 58(2–3),151–178.

[21] Bagnall, A., Ratanamahatana,C.A., Keogh,E., Lonardi,S., Janacek,G.A., 2006.Bit level representation for timeseries data mining with shape based similarity. Data Mining and Knowledge Discovery 13 (1), 11–40.

[22] Smyth, P., Keogh,E., 1997. Clustering and mode classification of engineering time series data. In Proceedings of the Third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.24–30.

[23] Shatkay, H., Zdonik,S., 1996. Approximate queries and representations for large data sequences. In Proceedings of the 12th IEEE International Conference on Data Engineering, pp.536–545.

[24] Ge, X.P., 1998. Pattern matching in financial time series data. University of California, Irvine, Final Project Report for. ICS 278.

[25] Perng, C.S., Wang,H., Zhang,R., Parker,D., 2000. Landmarks: a new model for similarity-based pattern querying in timeseries databases. In Proceedings of the 16th IEEE International Conference on Data Engineering, pp.33–42.

[26] Megalooikonomou, V., Li,G., Wang,Q., 2004. A Dimensionality reduction technique for efficient similarity analysis of timeseries databases. In Proceedings of the 13th ACM International Conference on Information and Knowledge Management, pp.160–161.

[27] Megalooikonomou, V., Wang,Q., Li,G., Faloutsos,C., 2005.A Multiresolution symbolic representation of timeseries .In Proceedings of the 21st IEEE International Conference on Data Engineering, pp.668–679.

[28] Moon, Y., Whang,K., Loh,W., 2001.Duality-based subsequence matching in time- series databases. In Proceedings of the 17th IEEE International Conference on Data Engineering, pp.263–272.

[29] Moon, Y.S., Whang,K.Y., Han,W.S., 2002.General match: a subsequence matching method in time-series databases based on generalized windows. In Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, pp.382–393.

[30] Loh, W.K., Moon,Y.S., Srivastava,J., 2010.Distortion-free predictive streaming time-series matching.InformationSciences180(8),1458–1476.

[31] Chung, F.L., Fu,T.C., Luk,R., Ng,V., 2001.Flexible timeseries pattern matching based on perceptually important points. In International Joint Conference on Artificial Intelligence Workshop on Learning from Temporal and Spatial Data, pp. 1–7.

[32] Chung,F.L., Fu,T.C., Ng,V., Luk,R., 8,2004. An evolutionary approach to pattern-based timeseries segmentation. IEEE Transactions on Evolutionary Computation, 471–489.

[33] van Wijk,J.J., vanSelow,E.R., 1999.Cluster and calendar based visualization of Time Series Data. In Proceedings of the IEEE Symposium on Information Visualization, pp.4–9.

[34] Schreck, Tekusova, Kohlhammer,T., FellnerD., J., 2007.Trajectory-based visual analysis of large financial timeseries data. ACM SIGKDD Explorations Newsletter, SpecialIssueonVisualAnalytics9(2),30–37.

[35] Weber, M., Alexa,M., Muller,W., 2001.Visualizing timeseries on spirals. In Proceedings of the IEEE Symposium on Information Visualization, pp.7–14.

[36] Hao, M.C., Dayal,U., Keim,D.A., Schreck,T., 2007.Multi-resolution techniques for visual exploration of large time-series data. In Proceedings of the Joint Eurographics—IEEE VGTC Symposium on Visualization, pp.27–34.

[37] Lin, J., Li,Y., 2009.Finding structural similarity in timeseries data using bag-of- patterns representation. In Proceedings of the 21st International Conference on Scientific and Statistical Database Management, pp.461–477.

[38] Lin, J., Keogh,E., Lonardi,S., 2005. Visualizing and discovering non-trivial patterns in large timeseries databases. Information Visualization 4 (2), 61–82.

[39] Lin,J., Keogh,E., Lonardi,S., Chiu,B., 2003. A Symbolic representation of timeseries , With implications for streaming algorithms. In Proceedings of the Eighth ACM SIGMOD International Conference on Management of Data Workshop on Research Issues in Data Mining and Knowledge Discovery, pp.2–11.

[40] Lin, J., Keogh,E., Lonardi,S., Patel,P., 2002. Finding motifs in timeseries .In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2nd Workshop on Temporal Data Mining, pp. 53–68.

[41] Lin, J., Keogh,E., Wei,L., Lonardi,S., 2007. Experiencing SAX: a novel symbolic representation of timeseries .Data Mining and Knowledge Discovery, 15(2), 107–144.

[42] Fu, T.C., Chung, F.L., Luk,R., Ng,C.M., 2008. Representing financial timeseries based on data point importance. Engineering Applications of Artificial Intelligence 21(2),277–300.

[43] Fu, T.C., Chung,F.L., Luk,R., Ng,C.M., 2007. Stock timeseries pattern matching: template-based vs. rule-based approaches. Engineering Applications of Artificial Intelligence20(3),347–364.

[44] Fu, T.C., Chung,F.L., Luk,R., Ng,V., 2001. Pattern discovery from stock timeseries using self-organizing maps. In Proceedings of the 7th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Temporal Data Mining, pp.27–37.

[45] Fu, T.C., Chung,F.L., Ng,C.M., 2006. Financial timeseries segmentation based on specialized binary tree representation. In Proceedings of the 2006 International Conference on Data Mining, pp.3–9.

[46] Fu, T.C., Chung,F.L., Tang,P.Y., Luk,R., Ng,C.M., 2005. Incremental stock time series data delivery and visualization. In Proceedings of the 14th ACM Conference on Information and Knowledge Management, pp.279–280.

[47] Chiu, B., Keogh,E., Lonardi,S., 2003. Probabilistic discovery of timeseries motifs. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.493–498.

[48] Tanaka,Y., Iwamoto,K., Uehara,K., 2005. Discovery of time-series motif from multi- Dimensional data based on MDL principle. Machine Learning, 58(2–3),269–300.

[49] Caraa-Valente, J.P., Lopez-Chavarrias,I., 2000. Discovering similar patterns in time series. In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data mining, pp.497–505.

[50] Lerner, A., Shasha,D., Wang,Z., Zhao,X., Zhu,Y., 2004. Fast algorithms for time series with applications to finance, physics, music, biology, and other suspects. In Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, pp.965–968.

[51] Chan, P., Mahoney,M., 2005.Modelingmultiple timeseries for anomaly detection. In Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 90–97.

[52] Kalpakis, K., Gada,D., Puttagunta,V., 2001. Distance measures for effective clustering of ARIMA time-series. In Proceedings of the IEEE International Conference on Data Mining,2001, pp.273–280.

[53] Panuccio, A., Bicego,M., Murino,V., 2002. A Hidden Markov Model-based approach to sequential data clustering. In the Joint International Association for Pattern Recognition Workshops on Structural, Syntactic and Statistical Pattern Recognition, pp.734–742.

[54] Xiong, Y., Yeung,D.Y., 2004. timeseries clustering with ARMA mixtures. Pattern Recognition 37 (8), 1675–1689.

[55] Oates, T., 1999. Identifying distinctive subsequences in multivariate timeseries by clustering. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.322–326.

[56] Oates, T., Firoiu,L., Cohen,P.R., 1999. Clustering timeseries with Hidden Markov Models and dynamic time warping. In Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Sequence Learning.

[57] Duan, J., Wang,W., Liu,B., Xue,Y., Zhou,H., Shi,B., 2005. Incorporating with recursive model training in timeseries clustering. In Proceedings of the Fifth International Conference on Computer and Information Technology, pp. 105–109.

[58] Geurts, P., 2001. Pattern extraction for timeseries classification. In Proceedings of the Fifth European Conference on Principles and Practice of Knowledge Discovery in Databases, pp.115–127.

[59] Zhang, H., Ho,T.B., Lin,M.S., 2004. A Non-parametric wavelet feature extractor for time-series classification. In Proceedings of the Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp.595–603.

[60] Zhang, T., Yue,D., Gu,Y., Yu,G., 2007. Boolean representation based data-adaptive correlation analysis over timeseries streams. In Proceedings of the 16th ACM Conference on Information and Knowledge Management, pp.203–212.

[61] Kadous, M.W., Sammut,C., 2005. Classification of multivariate timeseries and Structured data using constructive induction. Machine Learning, 58(2–3), 179–216

[62] Yang, K., Shahabi,C., 2005. A Multi level distance-based index structure for multivariate timeseries. In Proceedings of the 12th IEEE International Symposium on Temporal Representation and Reasoning, pp.65–73.

[63] Yang, K., Shahabi,C., 2005. On the stationarity of multivariate timeseries for correlation-based data analysis. In Proceedings of the Fifth IEEE International Conference on Data Mining, pp.805–808.

[64] Yang, K., Yoon,H., Shahabi,C., 2005. CLeVer: a feature subset selection technique for multivariate timeseries. In Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp.516–522.

[65] Povinelli, J., Feng,X., 1999. Datamining of multiple nonstationary timeseries. In Proceedings of Artificial Neural Networks in Engineering, pp.511–516.

[66] Rodriguez, J.J., Alonso,C.J., 2004. Interval and dynamic time warping-based decision trees. In Proceedings of the 2004 ACM Symposium on Applied Computing, pp.548–552.

[67] Kahveci, T., Singh,A.K., 2004. Optimizing similarity search of arbitrary length time series queries. IEEE Transactions on Knowledge and Data Engineering 16 (4), 418–433.

[68] Kahveci, T., Sing,A., Gurel,A., 2002. Similarity searching for multi-attribute sequences. In Proceedings of the 14th International Conference on Scientific and Statistical Database Management, pp.175–186.

[69] Gilbert, A.C., Kotidis,Y., Muthukrishnan,S., Strauss,M.J., 2001. Surfing wavelets on streams: one-pass summaries for approximate aggregate queries. In Proceedings of the 27th International Conference on Very Large Databases, pp.79–88.

[70] Moses, Diana. A survey of data mining algorithms used in cardiovascular disease diagnosis from multi-lead ECG data. Kuwait Journal of Science 42.2 (2015).

[71] Moses, Diana, and C. Deisy. "m-CADE: A mobile based cardiovascular abnormality detection engine using efficient multi-domain feature combinations." *Intelligent Data Analysis*20.3 (2016): 575-596.

[72] Deisy, C., Diana, M., Feature Selection for Nominal, Categorical and ECG Data, ISBN : 978-93-85977-78-7

[73] Gao, L., Wang,X.S., 2002. Continually evaluating similarity-based pattern queries on a Streaming timeseries .In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 370–381.

[74] Gao, L., Yao,Z., Wang,X.S., 2002. Evaluating continuous nearest neighbour queries for streaming timeseries via pre-fetching. In Proceedings of the 11th ACM International Conference on Information and Knowledge Management, pp. 485–492.

[75] Vlachos, M., Gunopulos,D., Kollios,G., 2002. Discovering similar multidimensional trajectories. In Proceedings of the 18th IEEE International Conference on Data Engineering, pp.673.