ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

A Survey on Time Series Data Mining

Kumar Vasimalla

Dept of Computer Science *(SMPS),* Central University of Kerala, India

ABSTRACT: To provide an overview this paper surveys and summarizes previous works done in the clustering, classification andsegmentation of time series data in various application domains. The basics of time series mining are presented, including measures to determine similarity/dissimilarity between two time series being compared, general purpose data mining algorithms commonly used in time series data mining, the criteria for evaluating the performance of the mining results and we hope that this review will serve as a stepping stone to researchers in advancing this area of research.

KEYWORDS: time-series,similarity,classification,clusteringand segmentation.

I. INTRODUCTION

A Time series is a set of observations each one being recorded at a specific time t.It is of two types, a discrete-time series is one in which the set of times at which observations are made is a discrete set, for example when the observations are made at fixed time intervals. Continuous time time-series are observed when observations are recorded continuously over some time interval. Major time-series-related tasks include query by content, anomalydetection, motif discovery, prediction, clustering, classification and segmentation. Time-series data mining unveils numerous facets of complexity. The most prominent problems are similarity measures,data representations and indexing methods. This paper reviewed some of the time-series data mining tasks.

The remaining part of this paper is organized as follows, Section 2 contains the definitions for terms used. The concept of similarity measure,and comparison of similarity measures reviewed in Section 3. The research work on time series classification and clustering and segmentation discussed in Sections 4. Whereas the conclusion will be made in Section 5.

II. DEFINITIONS

The purpose of this section is to provide a definition for the terms used throughoutthisarticle.
Definition 2.1.A *time-series T* is an ordered sequence of *n* real-valued variables $T = (t_1, \ldots, t_n), t_i \in R$. A time series is often the result of the observation of an underlying process in thecourse of which values are collected from measurements made at uniformly spacedeaseof use*time instants*and according to a given *sampling rate.* A time series can thus be defined as a set of contiguous time instants. The series can be *univariate* as in definition 2.1 or *multivariate* when several series simultaneously span multiple dimensions within the same time range.Time series can cover the full set of data provided by the observation of a process and may be of considerable length. In the case of streaming, they are semi-infinite astime instants continuously feed the series. It thus becomes interesting to consider onlythe *subsequences*of a series.

III. TIME SERIES SIMILARITY MEASURES

3.1 Euclidean Distances and LpNorms:

The oldest and simplest similarity measures for time series is the Euclidean distance(ED) measure. The restriction in ED is that both time series are of the same length *n*, and define the dissimilarity between series *C* and *Q* is *D*(*C,Q*) = *Lp*(*C,Q*), i.e. the distance between the two points measured by the *Lp*norm (when *p* = 2,it reduces to the familiar Euclidean distance).
Euclidean distance is the most widely used distance measure forsimilarity search [Agrawal *et al.*, 1993; Chan and Fu, 1999; Faloutsos*et al.*,1994]. However, one major disadvantage is that it is very brittle, it does not allowfor a situation

where two series are alike, but one has been “stretched”or “compressed” in the Y -axis. For example, a time series may fluctuate withsmall amplitude between 10 and 20, while another may fluctuate in a similarmanner with larger amplitude between 20 and 40. The Euclidean distance betweenthe two time series will be large. This problem can be dealt easilywith offset translation and amplitude scaling, which requires normalizing the series before applying the distance measure.In [Goldin and Kanellakis 1995], the authors describe a method where theseries are normalized in an effort to address the disadvantages of the Lpasa similarity measure. Figure 1 illustrates the idea more formally.

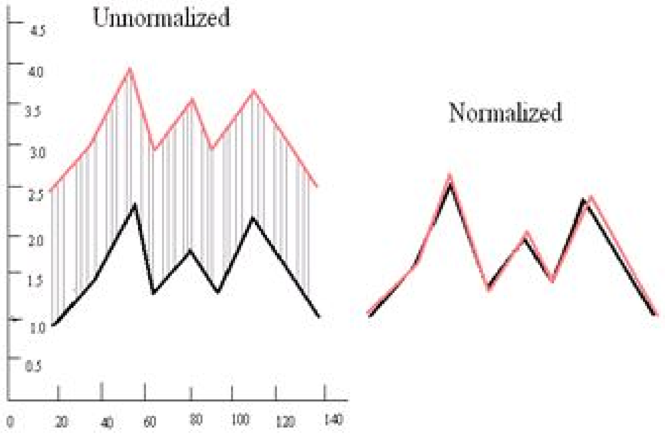

Figure 1 :A visual intuition of the necessity to normalize time series before measuring the distance between them. The two series Q and C appear to have approximately the sameshape, but have different offsets in Y-axis. The unnormalized data greatly overstate the subjective dissimilarity distance. Normalizing the data reveals the true similarity of the two time series

let μ(C) and σ(C) be the mean and standard deviation of sequence $T = (t_1, \ldots, t_n)$, $t_i \in R$.The series$T$is replaced by the normalized series T', Where $T_i' = T_i - \mu(T)/\sigma(T)$. Even after normalization, the Euclidean distance measure may still be unsuitablefor some time series domains since it does not allow for accelerationand deceleration along the time axis. For example, consider the two subjectivelyvery similar series shown in figure 2Even with normalization,the Euclidean distance will fail to detect thesimilarity between the two signals.This problem can generally be handled by Dynamic Time Warping(DTW) distancemeasure, which will be discussed in the next section.

3.2 Dynamic TimeWarping

It is often the case that the two series have approximately the same overall component shapes, but these shapes do not line up in X-axis. Figure 2 shows this with a simple example. In order to find the similarity between such series or as a pre-processing step before averaging them, we must “warp” the time axis of one (or both) series to achieve a better alignment. Dynamic Time Warping is a technique for effectively achieving this warping.

Dynamic time warping is an extensively used technique in speech recognition, and allows acceleration deceleration of signals along the time dimension, Although this dynamic programming technique is impressive in its abilityto discover the optimal of an exponential number alignments, a basic implementationruns in $O(mn)$ time. If a warping window w is specified, then the running time reduces to $O(nw)$, which is still too slow for most large scale application. In (Ratanamahatana and Keogh, 2004),the authors introduce a novel framework based on a learned warping window constraint to further improve the classification accuracy, as well as to speed up the DTW calculation by utilizing the lower bounding technique introduced in(Keogh, 2002).

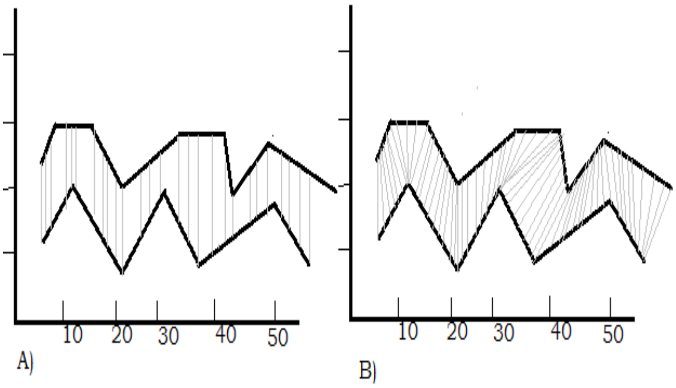

Figure 2:Note that while the series have an overall similar shape, they are not aligned in the time axis. Euclidean distance, whichassumes the i^{th}point on one sequence is aligned withi^{th}point on the other (A), will producea pessimistic dissimilarity measure. A nonlinear alignment (B) allows a more sophisticateddistance measure to be calculated

3.3 Longest Common Subsequence Similarity

The longest common subsequence(LCSS) similarity measure is a type of edit distance used in speech recognition and text pattern matching. The basic idea is to match two series by allowing some elements to be unmatched. The advantage of the LCSS method is that some elements may be unmatched or left out (e.g. outliers), where as in Euclidean and DTW, all elements from both sequences must be used, even the outliers. As was done with dynamic time warping, we give a recursive definition of the length of the longest common subsequence of C and Q. Let L (i, j) denote the longest common subsequences $\{c_1, \ldots, c_i\}$ and$\{q_1, \ldots, q_j\}$. L(i, j) may be recursively defined as shown in figure 3.

We define the dissimilarity between C and Q as$LCSS = m + n - 2l/m + n$. Where$l$ is the length of the longest common subsequence. Intuitively, this quantity determines the minimum (normalized) number of elements that should be removed from and inserted into C to transform C to Q. As with dynamic time warping, the LCSS measure can be computed by dynamic programming in O (mn) time. This can be improved to O ((n + m) w) time if a matching window of length w is specified (i.e. where |i −j| is allowed to be at most w).With time series data, the requirement that the corresponding elements in the common subsequence should match exactly is rather rigid. This problem is addressed by allowing some tolerance (say ε > 0) when comparing elements. Thus, two elements a and b are said to match if a(1 – ε) > b>a(1 + ε).

```
If $a_i = b_j$ then
$L(i,j) = 1+ L(i-1,j-1)$
else
$L(i,j) = \max\{D(i-1,j),D(I,j-1)\}$
```

Fig: 3 LCSS

Comparison of Distance Measures:

The selection of similarity measure depends on the type of data to be analysed and application to be developed on the data. If series are short and visual perception description available shape based methods are good, model based methods are suitable if application is targeting to a specific data set,if periodicity is the central subject of interest then feature-based methods are more appropriate.Even with these general recommendations and comparisons for the selection of an appropriate distance measure, the accuracy of the similarity chosen still has to be evaluated. The acuracy of distance measure is usually evaluated using 1-NN classifier framework. It has been shown by [Ding et al. 2008]. The table I summarizes the properties of various distance measures.

Table I: Comparison of Similarity measures

Similarity Measure	Time Complexity	Warp	Scale	Type
Dynamic Time Warping(DTW)	O(n^2)	Yes	No	Shape-based
Ecludian Distance(L_p Norms)	O(n)	No	No	Shape-based
LongestCommonSubSeq. (LCS	O(n)	Yes	No	Edit-based
Levenshtein	O(n^2)	No	No	Edit-based
Weighted Levenshtein	O(n^2)	No	No	Edit-based
LB-Keogh(DTW)	O(n)	Yes	No	Shape-based
Spatil Assembling(SpADe)	O(n^2)	Yes	Yes	Shape-based
Likelihood	O(n)	No	No	Feature-based
Auto correlation	O(nlogn)	No	No	Feature based
Vector Quantization	O(n^2)	Yes	No	Feature-baesd
Histogram	O(n)	No	No	Featur-based
Markov Chain(MC)	O(n)	No	No	Model-based
Hidden Markov Models(HMM)	O(n^2)	No	Yes	Model-based
Auto-Regresive(ARMA)	O(n^2)	No	No	Model-based
Kullback-Leibler	O(n)	No	No	Model-based

IV. TIME SERIES DATA MINING

4.1 Classification

As in classification, [Liao, 2005] concluded that all the algorithms designed for clustering time-series data either try to modify the existing static data algorithms to handle the sequential data, or modify the data itself for the existing algorithms to be able to handle it. His openionon dealing with time-series data directly is, find new similarity measures suitable for the time series data. Whereas those doing conversion on the sequential data either extract a feature-vector from it to be fed to the classifier (clustering algorithm is hiscase), or come out with a model for the data. [Keogh and Kasetty, 2003] limited their review to classification algorithms that rely on providing new similarity measures, while [Xing et al., 2010], on the other hand, categorized the classification algorithms into a similar categorization to those of [Liao, 2005]. Similarly, we are going to study the classification algorithms in the followingorder

- Distance-based classification
- Feature-based classification
- Model-based classification

Distance based classification:

Classification done based on distance between data elements is called distance based classification algorithms, example k-nearest neighbour (kNN). To apply conventional classification algorithms to time series data, new similarity measures are required for sequential data. [Xing et al.,2010]argues that distance measures decides the acuracy of classification algorithm.[Keogh and Kasetty, 2003, Ratanamahatana and Keogh, 2004a] emphasised on its sensitivity to distortion in time. distortion is also non-linear, hence linear transformation will not be sufficient. elastic similarity measuressuch as Dynamic time warping distance (DTW) were needed to solve this problem.[Ratanamahatana and Keogh, 2004a] explained DTW as a non-linear mapping between two series where the distance between them is minimized. Although many researchers [Aach and Church, 2001, Bar-Joseph et al., 2002, Yi et al.,1998] agreed on the superiority of DTW over Euclidean distance, its computational inefficiently is limiting itsadoption [Ratanamahatana and Keogh, 2004b]. DTW is calculated using dynamic programming, hence has a quadratic time complexity (O(n^2)), some researchestried to exploit this fact, in addition to the constrains, in order to speed up the DTW calculations [Xi et al., 2006].

[Durbin et al., 1998] highlighted that algorithms (such Needleman-Wunsch [Needleman et al., 1970] and Smith-Waterman [Smith and Waterman, 1981]) are calculated using dynamic programming, hence their complexity is O(n^2). Hence, as noted by [Vinga and Almeida, 2003], more optimum algorithms such as BLAST [Altschul et al., 1990] and

FASTA [Pearson et al., 1990] were presented later on. The newer algorithms use heuristic approaches, which means that although they are faster in comparing series [Tatusova and Madden, 2006], they do not guarantee in finding the optimal score [Durbin et al., 1998]. Additionally, BLAST 2.0 [Tatusova and Madden, 2006] is a tool that utilizes BLAST engine for pairwise sequence comparison, yet it is proposed as an alternative when comparing two series that are already known to be homologous.As mentioned earlier, sequential data can be multivariate. [Yang and Shahabi, 2004] noticed that breaking multivariate time series (MTS) into separate series and processing each one on its own result in overlooking the correlation between those variables. They presented a newer distance-measurement algorithm, Eros (Extended Frobenius norm), in order to deal with MTS.

Feature Based Classification:

Feature based classification algorithms, do their classificationbasedon feature-set, example ANN and Decision Trees. To apply feature based classification to time series data first we have to transform sequential data into feature set. [Xing et al., 2010].The choice of the appropriate features is the hardest part of this process, and as mentioned by[Eads et al., 2005], there is always trade-off between doing this process manually by domain experts or having it automated but less accurate in many cases. Patterns and wavelet decomposition, as we will see now, are ways for extracting features from sequential data.

[Ye and Keogh, 2009] noticed that algorithms that try to identify tree-leaves based on theirshapes are mislead by the deformation in their shapes due to insects eating parts of them.Instead of relying on the whole shape of the leaves (global features), they selected local features(patterns) that particularly discriminates the leaves from different trees. They converted theshape data into a sequential one. The aim is to find sub series, or shapelets as they called them that are discriminating between classes. To determine which subseries are to bechosen, they ordered all series according to their (Euclidean) distance from all possibleshapelets. Then they started to search for a mid-point that divides member-series of eachclass. Having a discriminative approach [Leslie et al., 2002], i.e. binary decisions are taken totell whether a new sequence belongs to a certain class or not, [Ye and Keogh, 2009] had to usea decision trees in their classifier. The more classes we have the more branches and split pointshas the tree. Similarly, [Ji et al., 2005] introduced a pattern-extraction algorithm called Minimal Distinguishing Subsequence (MDS). However, MDS allow for gaps with in the sub-series, which makesit more suitable to classifying biological series as mentioned earlier.

Another feature-extraction technique is to transform the time-series data into the frequency domain, where data dimensionality can be reduced. [Yang and Shahabi,2004] listed DFT (DiscreteFourier Transform), DWT (Discrete Wavelet Transform) and SVD (Singular Value Decomposition) as examples here. However, [Li et al., 2005] notes that DWT is more common in classification since it preserves both time and frequency characteristics, whereas DFT provides thefrequency characteristics only. Such transformation also solves a problem discussed earlier,where we need to study both local and broad trends within the sequential data [Aggarwal, 2002].

DWT transforms the data into different frequency components [Daubechies et al., 1992]. Thecomponents with higher order coefficients reflect the global trends of the data, while the ones with lower order coefficients reflect the local trends in it [Aggarwal, 2002].Kernel methods (KM) are also good in feature extraction, additionally, they can deal with symbol-series with diffierent lengths [Watkins, 1999]. Although [Joachims, 1998] was dealing with text data as a bag of words rather than sequential data, he highlighted the abilityof kernel methods to deal with textual data regardless of its huge number of features, normally more than 10k. He was using Support Vector Machine in particular, which is one of thekernel methods. KM calculates the inner product of the input vectors in a high dimensionalspace [Lodhi et al., 2002]. By doing so, linear decision boundaries can be drawn between theclasses [Leslie et al., 2002]. Unlike [Joachims, 1998], [Lodhi et al., 2002] used KM to classifytext as sequential data. Like alignment-based distance measures, kernel methods are widelyused in biological series classification [Liao and Noble, 2003, Zavaljevski et al., 2002].

Model based classification:

Model-based methods works by dividing the data into test data and training data, using the traing data construct a model and train the training dataset on the model to classify the training data[Liao, 2005]. He divided the models used in classification into statistical and neural network ones. According to [Rabiner, 1989], the statistical models such as Gaussian, Poisson, Markov and HiddenMarkov Models, are constructed.[Laxman and Sastry, 2006, Dunham, 2002], on the other hand, divided models into predictivemodels that tries to predict unavailable values of the data using the existing one, and descriptivemodelsthat tries to find patterns and relationships in the data,especially Markov models which are used a lot in sequence classification applications [Laxman and Sastry, 2006].

Hidden Markov Model (HMM) is defined by [Baldi et al., 1994]. [Birney, 2001] argues that HMM is more successful in biological seriesclassifications,compared to Neural Networks, since it can deal with variable-length series, while the other technique require fixed-length inputs. [Rabiner, 1989], on the other hand, pinpointed some ofHMM general limitations, [Graves et al., 2006] criticize the assumption of statesprobability independence, adding that HMM requires prior domain-specific knowledge to choosethe input features.Generally, artificial neural networks (ANN) are very close to statistical models [Ruck et al., 1990]. [Giles et al., 2001] defines recurrent neural networks (RNN) as special type of ANN, where thereis a feedback connection in the network to keep track of its internal state when dealing withnew inputs. RNN is suitable for sequential data since, according to [Giles et al., 2001], RNN iscapable of modelling the temporal nature of the sequence. Also, [Graves et al., 2006] stated thatin contrast to HMM, RNN does not require knowledge of the data. He also claimed that RNNis immune to temporal noise. Nevertheless, as seen earlier, they require fixed-length inputs.

4.2 CLUSTERING

Time series clustering algorithms form clusters, based on the type of time series data, distinctions can be made as to whether the data are discrete-valued or real-valued, uniformly or non-uniformly sampled, univariate or multivariate, and whether data series are of equal or unequal length. Non-uniformly sampled data must be converted into uniformed data before clustering operations can be performed. Various algorithms have been developed to clusterdifferenttypesof time series data.This paper groups previously developed time series clustering methods into three major categories depending upon whether they work directly with raw data, indirectly with features extracted from the raw data, or indirectly with models built from the raw data.

- Raw-data-based clustering
- Feature-based clustering
- Model-based clustering

Raw-data-based clustering:

Methods that work with raw data, either in the time or frequency domain, are placed into this category. The two time series beingcompared are normally sampled at the same interval, but their length (or number of time points) might or might not be the same. For clustering multivariate time varying data, [Kosmelj and Batagelj 1990]modified the relocation clustering procedurethat was originally developed for static data. To form a specified number of clusters, the best clustering among all the possible clusterings is the one with the minimum generalizedWard criterion function.[Kumar et al. 2002] proposed a *distance function based on the assumed independent Gaussian models of data errors* and used a *hierarchical clustering* method to group seasonality series into a desirable number of clusters. For the analysis of dynamic biomedical image time seriesdata, [Wismuller et al. 2002] showed that deterministic annealing by the minimal free energy vector quantization (VQ) could be effective. [Moller-Levet et al. 2003]proposed*short time series (STS) distance* to measurethe similarity in shape formed by the relative change ofamplitude and the corresponding temporal information ofuneven sampling intervals.

To group multivariate vector series of earthquakes andmining explosions, [Kakizawa et al. 1998] applied hierarchicalclustering as well as *k-means clustering*. [Shumway] investigated the clustering of nonstationarytime series by applying locally stationary versionsof*Kullback–Leibler discrimination information measures*that give optimal time–frequency statistics for measuringthe discrepancy between two non-stationary time series. [Policker and Geva 2000] modeled non-stationary time serieswith a time varying mixture of stationary sources, comparableto the continuous hidden Markov model. [Liao 2005] developed a two-step procedure for clustering multivariate time series of equal or unequal length. The firststep applies the *k-means* or *fuzzy c-means* clustering algorithm to time stripped data in order to convert multivariate real-valued time series into univariate discrete-valued time series. The second step employs the *k*-means or FCM algorithm again to group the converted univariate time series.

Feature-based clustering:

It is always not possible to work directly with the raw data that are highly noisy. Several feature-based clustering methods have been proposed to address these concerns. Though most feature extraction methods are generic in nature, the extracted features are usually application dependent. That is, one set of features that work well on one application might not be relevant to another. Some studies even take another feature selection step to further reduce the number of feature dimensions after feature extraction.[WilponandRabiner 1985] modified the standard k-means clustering algorithm for the recognition of isolated words. To measure the distance between two spoken word patterns, *a symmetric distance measure* was defined based on the Itakura distance for measuring the distance between two frames. The proposed modified k-means (MKM) clustering algorithm was shown to out perform the well established unsupervised without averaging (UWA) clustering algorithm at that time.[Shaw and King] clustered time series indirectly by applying two hierarchical clustering algorithms, the *Ward's minimum variance algorithm* and the *single linkage algorithm*, to normalized spectra (normalized by the amplitude of the largest peak). The spectra were constructed from the original time series with the means adjusted to zero. [Goutte et al. 2001] Clustered fMRI time series (P slices of images) in groups of voxels with similar activations using two algorithms: *k-means* and *Ward' s hierarchical clustering*. The *cross-correlation function* between the fMRI activation and the paradigm (or stimulus) was used as the feature space[Fu et al. 2001] described the use of self-organizing maps for grouping data sequences segmented from the numerical time series using a continuous sliding window with the aim to discover similar temporal patterns dispersed along the time series.

Model-based clustering:

This class of approaches considers that each time series is generated by some kind of model or by a mixture of underlying probability distributions. Time series are considered similar when the models characterizing individual series or the remaining residuals after fitting the model are similar.

For clustering or choosing from a set of dynamic structures (specifically the class of ARIMA invertible models), [Piccolo 1990] introduced the Euclidean distance between their corresponding autoregressive expansions as the metric. [Baragona] evaluated three meta-heuristic methods for partitioning a set of time series into clusters. Motivated by questions raised in the context of musical performance theory, [Beran and Mazzola] defined hierarchical smoothing models (or HISMOOTH models) to understand the relationship between the symbolic structure of a music score and its performance, with each represented by a time series. [Maharaj] developed an agglomerative hierarchical clustering procedure that is based on the p-value of a test of hypothesis applied to every pair of given stationary time series. [Ramoni et al] presented BCD: a Bayesian algorithm for clustering by dynamics. [Kalpakis et al.]Studied the clustering of ARIMA time-series, by using the Euclidean distance between the Linear Predictive Coding cepstra of two time-series as their dissimilarity measure.[Xiong and Yeung 2002] proposed a model-based method for clustering univariate ARIMA series. Assuming the Gaussian mixture model for speaker verification, [Tran and Wagner] proposed a fuzzy c-means clustering-based normalization method to find a better score to be compared with a given threshold for accepting or rejecting a claimed speaker.

4.3 SEGMENTATION

The segmentation problem can be framed in several ways.

- Given a time series T, produce the best representation using only K segments.

- GivenatimeseriesT,producethebestrepresentationsuchthatthe maximum error for any segment does not exceed some user-specified threshold, max-error.

- Given a time series T, produce the best representation such that the combined error of all segments is less than some user-specified threshold, total-max-error.

All algorithms can support all these specifications. Segmentation algorithms can also be classified as batch or

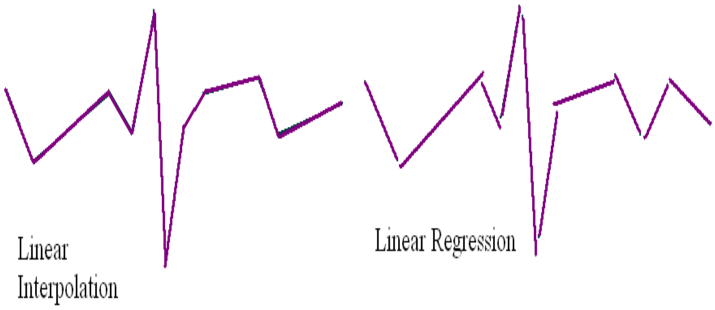

Fig 4: Two 10-segment approximations of electrocardiogram data. The approximation created using linear interpolation has a smooth aesthetically appealing appearance because all the endpoints of the segments are aligned. Linear regression, in contrast, produces a slightly disjointed appearance but a tighter approximation in terms of residual error.

online. This is an important distinction because many data mining problems are inherently dynamic [Vullings et al. (1997), Koski et al. (1995)]. Data mining researchers, who needed to produce a piecewise linear approximation, have typically either independently rediscovered an algorithm or used an approach suggested in related literature. For example, from the fields of cartography or computer graphics [Douglas and Peucker (1973), Heckbert and Garland (1997), Ramer (1972)].Here, we review the three major segmentation approaches in the literature and provide an extensive empirical evaluation on a very heterogeneous collection of datasets from finance, medicine, manufacturing and science. The major result of these experiments is that only online algorithm in the literature produces very poor approximations of the data, and that the only algorithm that consistently produces high quality results and scales linearly in the size of the data is a batch algorithm. The new online algorithm that scales linearly in the size of the data set, is online, and produces high quality approximations is SWAB(Sliding Window and Bottom-Up) [EamonnKeogh,Selina Chu, David Hart, and Michael Pazzani].Although appearing under different names and with slightly different implementation details, most time series segmentation algorithms can be grouped into one of the following three categories.

•*Sliding Windows:* A segment is grown until it exceeds some error bound. The process repeats with the next data point not included in the newly approximated segment.

• *Top-Down:* The time series is recursively partitioned until some stopping criteria is met.

• *Bottom-Up:* Starting from the finest possible approximation, segments are merged until some stopping criteria is met.

Algorithm/ Feature	Online	Time complexity	User can specify
Top-Down	No	$O(n^2K)$	E, ME, K
Bottom-Up	No	O(Ln)	E, ME, K
Sliding Window	Yes	O(Ln)	E

n-Number of data points, L- Average segment length,E-Maximum error for a given segment, ME- Maximum error for a given segment for entire time series, K- Number of segments

Table II: Comparison of Segmentation Algorithms

Given that we are going to approximate a time series with straight lines, there are two ways to find the approximating line.

- Linear Interpolation: Here the approximating line for the subsequence T[a : b] is simply the line connecting T_a and T_b. This can be obtained in constant time.
- Linear Regression: Here the approximating line for the subsequence T[a: b] is taken to be the best fitting line in the least squares sense. This can be obtained in time linear in the length of segment.

The two techniques are illustrated in figure 4. Linear interpolation tends to closely align the endpoint of consecutive segments, giving the piece-wise approximation a “smooth” look. In contrast, piecewise linear regression can produce a very disjointed look on some datasets. The aesthetic superiority of linear interpolation, together with its low computational complexity has made it the technique of choice in computer graphic applications [Heckbert and Garland (1997)]. However, the quality of the approximating line, in terms of Euclidean distance, is generally inferior to the regression approach. All segmentation algorithms also need some method to evaluate the quality of fit for a potential segment. A measure commonly used in conjunction with linear regression is the sum of squares, or the residual error. Taking all the vertical differences between the best-fit line and the actual data points, squaring them and then summing them together calculate this. Another commonly used measure of goodness of fit is the distance between the best-fit line and the data point furthest away in the vertical direction (i.e. the L∞ norm between the line and the data)in addition to the time complexity there are other features a practitioner might consider when choosing an algorithm as shown in table II.

First there is the question of the comparison of major segmentation algorithms.Whether the algorithm is online or batch. Secondly, there is the question of how the user can specify the quality of desired approximation. With trivial modifications the Bottom-Up algorithm allows the user to specify the desired value of *K*, the maximum error per segment, or total error of the approximation. A (non-recursive) implementation of Top-Down can also be made to support all three options. However Sliding Window only allows the maximum error per segment to be specified.

V.CONCLUSION

We have reviewed some major tasks in time-series data mining. Since time-series data are typically very large, discovering knowledge from these massive data becomes a challenge, which leads to enormous research challenges. The similarity measure is very important part of time series data mining, which decides the accuracy of data mining task. We review some of the important works of time series classification, clustering and segmentation. We would like to emphasize that the key step in any successful data mining endeavor always lies in choosing right representation of data and similarity measure for the task at hand.

REFERENCES

1. Chan, K., Fu, A.W. 1999Efficient time series matching by wavelets. Proceedings of 15th IEEE International Conference on Data Engineering; 1999 Mar 23-26, Sydney, Australia, pp. 126-133.
2. Agrawal, R., Faloutsos, C 1993, Efficient Similarity Search in Sequence Data bases. International Conference on Foundations of Data Organization (FODO), 1993.
3. GOLDIN, D. AND KANELLAKIS, P. 1995. On similarity queries for time-series data: Constraint specificationand implementation. In *Proceedings of the Principles and Practice of Constraint Programming (CP95)*.Springer, 137–153.
4. Keogh, E., Lonardi, S., Ratanamahatana, C.A. 2004] Towards Parameter-Free Data Mining. Proceedings of 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2004 Aug 22-25; Seattle, WA.
5. Keogh, E. and Kasetty, S. 2002] On the Need for Time Series Data Mining Bench- marks: A Survey and Empirical Demonstration. In the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2002 Jul 23 – 26; Edmonton, Alberta, Canada, pp 102-111.
6. DING, H., TRAJCEVSKI, G., SCHEUERMANN, P., WANG, X., AND KEOGH, E. 2008.] Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proc. VLDB Endowm. 1,* 2,1542–1552.
7. T. Warren Liao 2005,Clustering of time series data—a survey*Industrial & Manufacturing Systems Engineering Department, Louisiana State University, 3128 CEBA, Baton Rouge, LA 70803, USA* Received 16 September 2003; received in revised form 21 June 2004; accepted 7 January 2005 Elsevier.
8. K. Kosmelj, V. Batagelj 1990, Cross-sectional approach for clustering time varying data, J. Classification 7 (1990) 99–109.
9. M. Kumar, N.R. Patel, J. Woo 2002, Clustering seasonality patterns in the presence of errors, Proceedings of KDD ’02, Edmonton, Alberta, Canada.
10. Wismuller, O. Lange, D.R. Dersch, G.L. Leinsinger, K. Hahn, B. Putz, D. Auer 2002], Cluster analysis of biomedical image time series, Int. J. Comput. Vision 46 (2) (2002) 103–128.
11. C.S. Moller-Levet, F. Klawonn, K.H. Cho, O. Wolkenhauer 2003], Fuzzy clustering of short time series and unevenly distributed sampling points, Proceedings of the 5th International Symposium on Intelligent Data Analysis, Berlin, Germany, August 28–30, 2003.
12. Y. Kakizawa, R.H. Shumway, N. Taniguchi 1998, Discrimination and clustering for multivariate time series, J. Amer. (441) (1998) 328–340.
13. S. Policker, A.B. Geva 2000, Non-stationary time series analysis by temporal clustering, IEEE Trans. Syst. Man Cybernet.-B. 30 (2) (2000) 339–343.
14. J.G. Wilpon, L.R. Rabiner 1985, Modified k-means clustering algorithm for use in isolated word recognition, IEEE Trans. Acoust. Speech Signal Process. 33 (3) (1985) 587–594.
15. C. Goutte, L.K. Hansen, M.G. Liptrot, E. Rostrup 2001, Feature- space clustering for fMRI meta-analysis, Hum. Brain Mapping 13 (2001) 165–183.
16. T.C. Fu, F.L. Chung, V. Ng, R. Luk 2001, Pattern discovery from stock time series using self-organizing maps, KDD 2001 Workshop on

Temporal Data Mining, August 26–29, San Francisco, 2001, pp. 27–37.

17. D. Piccolo 1990, A distance measure for classifying ARMA models, J. Time Ser. Anal. 11 (2) (1990) 153–163.
18. Y. Xiong, D.Y. Yeung 2002, Mixtures of ARMA models for model-based time series clustering, Proceedings of the IEEE International Conference on Data Mining, Maebaghi City, Japan, 9–12 December, 2002.
19. Vullings, H.J L.M., Verhaegen, M.H.G., and Verbruggen H.B. (1997). ECG Segmentation Using Time-Warping. Proceedings of the 2nd International Symposium on Intelligent Data Analysis, pp. 275–286.
20. Douglas, D.H. and Peucker, T.K. (1973). Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. Canadian Cartographer, 10(2) December, pp. 112–122.
21. Xing, Z., Pei, J., and Keogh, E. (2010). A brief survey on sequence classification. ACM SIGKDD Explorations Newsletter, 12(1):40-48.
22. Ye, L. and Keogh, E. (2009)]. Time series shapelets: a new primitive fordata mining. In Proceedings of the 15th ACM SIGKDD international conference on Knowledgediscovery and data mining, pages 947{956. ACM.
23. Leslie, C., Eskin, E., and Noble, W. (2002). The spectrum kernel: A string kernel for svm protein classification. In Proceedings of the pacific symposium on biocomputing,volume 7, pages 566{575. Hawaii, USA.
24. Yang, K. and Shahabi, C. (2004). A pca-based similarity measurefor multivariate time series. In ACM International Workshop On Multimedia Databases:Proceedings of the 2nd ACM international workshop on Multimedia databases, volume 13, pages 65-74.
25. Aach and Church, 2001,Aach, J. and Church, G. (2001). Aligning gene expression time serieswith time warping algorithms. Bioinformatics, 17(6):495-508.