Engineering Applications of Artificial Intelligence 24 (2011) 164–181

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A review on time series data mining

Tak-chung Fu

Department of Computing, Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 19 February 2008
Received in revised form
14 March 2010
Accepted 4 September 2010

Keywords:
Time series data mining
Representation
Similarity measure
Segmentation
Visualization

ABSTRACT

Time series is an important class of temporal data objects and it can be easily obtained from scientific and financial applications. A time series is a collection of observations made chronologically. The nature of time series data includes: large in data size, high dimensionality and necessary to update continuously. Moreover time series data, which is characterized by its numerical and continuous nature, is always considered as a whole instead of individual numerical field. The increasing use of time series data has initiated a great deal of research and development attempts in the field of data mining. The abundant research on time series data mining in the last decade could hamper the entry of interested researchers, due to its complexity. In this paper, a comprehensive revision on the existing time series data mining research is given. They are generally categorized into representation and indexing, similarity measure, segmentation, visualization and mining. Moreover state-of-the-art research issues are also highlighted. The primary objective of this paper is to serve as a glossary for interested researchers to have an overall picture on the current time series data mining development and identify their potential research direction to further investigation.

1. Introduction

Recently, the increasing use of temporal data, in particular time series data, has initiated various research and development attempts in the field of data mining. Time series is an important class of temporal data objects, and it can be easily obtained from scientific and financial applications (e.g. electrocardiogram (ECG), daily temperature, weekly sales totals, and prices of mutual funds and stocks). A time series is a collection of observations made chronologically. The nature of time series data includes: large in data size, high dimensionality and update continuously. Moreover time series data, which is characterized by its numerical and continuous nature, is always considered as a whole instead of individual numerical field. Therefore, unlike traditional databases where similarity search is exact match based, similarity search in time series data is typically carried out in an approximate manner.

There are various kinds of time series data related research, for example, finding similar time series (Agrawal et al., 1993a; Berndt and Clifford, 1996; Chan and Fu, 1999), subsequence searching in time series (Faloutsos et al., 1994), dimensionality reduction (Keogh, 1997b; Keogh et al., 2000) and segmentation (Abonyi et al., 2005). Those researches have been studied in considerable detail by both database and pattern recognition communities for different domains of time series data (Keogh and Kasetty, 2002).

In the context of time series data mining, the fundamental problem is how to represent the time series data. One of the common approaches is transforming the time series to another domain for dimensionality reduction followed by an indexing mechanism. Moreover similarity measure between time series or time series subsequences and segmentation are two core tasks for various time series mining tasks. Based on the time series representation, different mining tasks can be found in the literature and they can be roughly classified into four fields: pattern discovery and clustering, classification, rule discovery and summarization. Some of the research concentrates on one of these fields, while the others may focus on more than one of the above processes. In this paper, a comprehensive review on the existing time series data mining research is given. Three state-of-the-art time series data mining issues, streaming, multi-attribute time series data and privacy are also briefly introduced.

The remaining part of this paper is organized as follows: Section 2 contains a discussion of time series representation and indexing. The concept of similarity measure, which includes both whole time series and subsequence matching, based on the raw time series data or the transformed domain will be reviewed in Section 3. The research work on time series segmentation and visualization will be discussed in Sections 4 and 5, respectively. In Section 6, vary time series data mining tasks and recent time series data mining directions will be reviewed, whereas the conclusion will be made in Section 7.

2. Time series representation and indexing

One of the major reasons for time series representation is to reduce the dimension (i.e. the number of data point) of the

E-mail addresses: cstcfu@comp.polyu.edu.hk, cstcfu@gmail.com

0952-1976/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2010.09.007

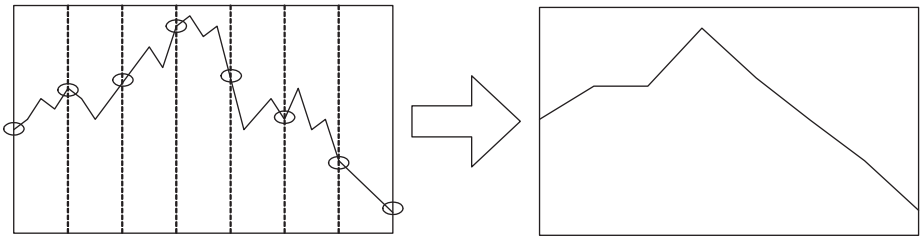

Fig. 1. Time series dimensionality reduction by sampling. The time series on the left is sampled regularly (denoted by dotted lines) and displayed on the right with a large distortion.

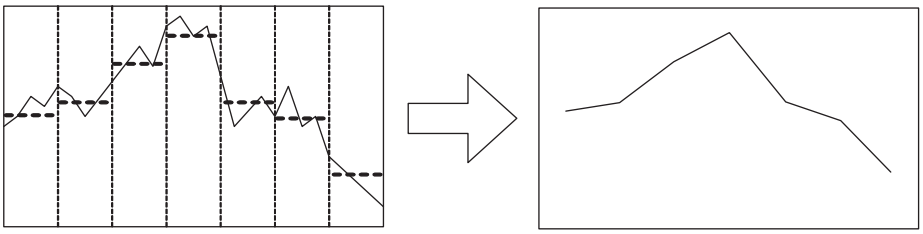

Fig. 2. Time series dimensionality reduction by PAA. The horizontal dotted lines show the mean of each segment.

original data. The simplest method perhaps is sampling (Astrom, 1969). In this method, a rate of m/n is used, where m is the length of a time series P and n is the dimension after dimensionality reduction (Fig. 1). However, the sampling method has the drawback of distorting the shape of sampled/compressed time series, if the sampling rate is too low.

An enhanced method is to use the average (mean) value of each segment to represent the corresponding set of data points. Again, with time series $P=(p_1,\ldots,p_m)$ and n is the dimension after dimensionality reduction, the "compressed" time series $\hat{P}=(\hat{p}_1,\ldots,\hat{p}_n)$ can be obtained by

$$\hat{p}_k=\frac{1}{e_k-s_k+1}\sum_{i=s_k}^{e_k}p_i \tag{1}$$

where s_k and e_k denote the starting and ending data points of the kth segment in the time series P, respectively (Fig. 2). That is, using the segmented means to represent the time series (Yi and Faloutsos, 2000). This method is also called piecewise aggregate approximation (PAA) by Keogh et al. (2000).[1] Keogh et al. (2001a) propose an extended version called an adaptive piecewise constant approximation (APCA), in which the length of each segment is not fixed, but adaptive to the shape of the series. A signature technique is proposed by Faloutsos et al. (1997) with similar ideas. Besides using the mean to represent each segment, other methods are proposed. For example, Lee et al. (2003) propose to use the segmented sum of variation (SSV) to represent each segment of the time series. Furthermore, a bit level approximation is proposed by Ratanamahatana et al. (2005) and Bagnall et al. (2006), which uses a bit to represent each data point.

To reduce the dimension of time series data, another approach is to approximate a time series with straight lines. Two major categories are involved. The first one is linear interpolation. A common method is using piecewise linear representation (PLR)[2] (Keogh, 1997b; Keogh and Smyth, 1997; Smyth and Keogh, 1997). The approximating line for the subsequence $P(p_i,\ldots,p_j)$ is simply the line connecting the data points p_i and p_j. It tends to closely align the endpoint of consecutive segments, giving the piecewise approximation with connected lines. PLR is a bottom-up algorithm. It begins with creating a fine approximation of the time series, so that $m/2$ segments are used to approximate the m length time series and iteratively merges the lowest cost pair of segments, until it meets the required number of segment. When the pair of adjacent segments S_i and S_{i+1} are merged, the cost of merging the new segment with its right neighbor and the cost of merging the S_i+1 segment with its new larger neighbor is calculated. Ge (1998) extends PLR to hierarchical structure. Furthermore, Keogh and Pazzani enhance PLR by considering weights of the segments (Keogh and Pazzani, 1998) and relevance feedback from the user (Keogh and Pazzani, 1999). The second approach is linear regression, which represents the subsequences with the best fitting lines (Shatkay and Zdonik, 1996).

Furthermore, reducing the dimension by preserving the salient points is a promising method. These points are called as perceptually important points (PIP). The PIP identification process is first introduced by Chung et al. (2001) and used for pattern matching of technical (analysis) patterns in financial applications. With the time series P, there are n data points: P_1, P_2 …, P_n. All the data points in P can be reordered by its importance by going through the PIP identification process. The first data point P_1 and the last data point P_n in the time series are the first and two PIPs, respectively. The next PIP that is found will be the point in P with maximum distance to the first two PIPs. The fourth PIP that is found will be the point in P with maximum vertical distance to the line joining its two adjacent PIPs, either in between the first and second PIPs or in between the second and the last PIPs. The PIP location process continues until all the points in P are attached to a reordered list L or the required number of PIPs is reached (i.e. reduced to the required dimension). Seven PIPs are identified in from the sample time series in Fig. 3. Detailed treatment can be found in Fu et al. (2008c).

The idea is similar to a technique proposed about 30 years ago for reducing the number of points required to represent a line by Douglas and Peucker (1973) (see also Hershberger and Snoeyink, 1992). Perng et al. (2000) use a landmark model to identify the important points in the time series for similarity measure. Man and Wong (2001) propose a lattice structure to represent the identified peaks and troughs (called control points) in the time series. Pratt and Fink (2002) and Fink et al. (2003) define extrema as minima and maxima in a time series and compress the time

[1] This method is called piecewise constant approximation originally (Keogh and Pazzani, 2000a).

[2] It is also called piecewise linear approximation (PLA).

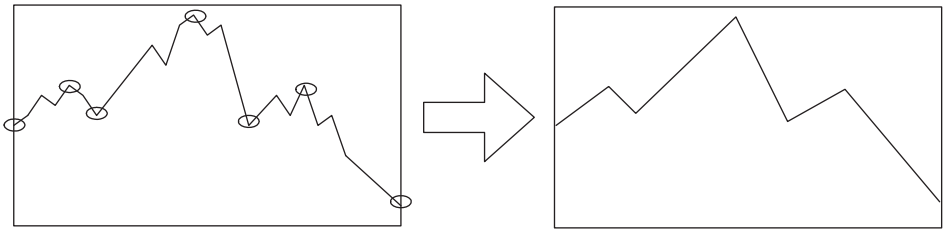

Fig. 3. Time series compression by data point importance. The time series on the left is represented by seven PIPs on the right.

series by selecting only certain important extrema and dropping the other points. The idea is to discard minor fluctuations and keep major minima and maxima. The compression is controlled by the compression ratio with parameter R, which is always greater than one; an increase of R leads to the selection of fewer points. That is, given indices i and j, where $i \le x \le j$, a point p_x of a series P is an important minimum if p_x is the minimum among $p_i, \ldots, p_j$, and $p_i/p_x \ge R$ and $p_j/p_x \ge R$. Similarly, p_x is an important maximum if p_x is the maximum among $p_i, \ldots, p_j$ and $p_x/p_i \ge R$ and $p_x/p_j \ge R$. This algorithm takes linear time and constant memory. It outputs the values and indices of all important points, as well as the first and last point of the series. This algorithm can also process new points as they arrive, without storing the original series. It identifies important points based on local information of each segment (subsequence) of time series. Recently, a critical point model (CPM) (Bao, 2008) and a high-level representation based on a sequence of critical points (Bao and Yang, 2008) are proposed for financial data analysis. On the other hand, special points are introduced to restrict the error on PLR (Jia et al., 2008). Key points are suggested to represent time series in (Leng et al., 2009) for an anomaly detection.

Another common family of time series representation approaches converts the numeric time series to symbolic form. That is, first discretizing the time series into segments, then converting each segment into a symbol (Yang and Zhao, 1998; Yang et al., 1999; Motoyoshi et al., 2002; Aref et al., 2004). Lin et al. (2003; 2007) propose a method called symbolic aggregate approximation (SAX) to convert the result from PAA to symbol string. The distribution space (y-axis) is divided into equiprobable regions. Each region is represented by a symbol and each segment can then be mapped into a symbol corresponding to the region in which it resides. The transformed time series $\hat{P}$ using PAA is finally converted to a symbol string $SS(s_1, \ldots, s_W)$. In between, two parameters must be specified for the conversion. They are the length of subsequence w and alphabet size A (number of symbols used). Besides using the means of the segments to build the alphabets, another method uses the volatility change to build the alphabets. Jonsson and Badal (1997) use the "Shape Description Alphabet (SDA)". Example symbols like highly increasing transition, stable transition, and slightly decreasing transition are adopted. Qu et al. (1998) use gradient alphabets like upward, flat and download as symbols. Huang and Yu (1999) suggest transforming the time series to symbol string, using change ratio between contiguous data points.

Megalooikonomou et al. (2004) propose to represent each segment by a codeword from a codebook of key-sequences. This work has extended to multi-resolution consideration (Megalooikonomou et al., 2005). Morchen and Ultsch (2005) propose an unsupervised discretization process based on quality score and persisting states. Instead of ignoring the temporal order of values like many other methods, the Persist algorithm incorporates temporal information.

Furthermore, subsequence clustering is a common method to generate the symbols (Das et al., 1998; Li et al., 2000a; Hugueney and Meunier, 2001; Hebrail and Hugueney, 2001). A multiple abstraction level mining (MALM) approach is proposed by Li et al. (1998), which is based on the symbolic form of the time series. The symbols in this paper are determined by clustering the features of each segment, such as regression coefficients, mean square error and higher order statistics based on the histogram of the regression residuals.

Most of the methods described so far are representing time series in time domain directly. Representing time series in the transformation domain is another large family of approaches. One of the popular transformation techniques in time series data mining is the discrete Fourier transforms (DFT), since first being proposed for use in this context by Agrawal et al. (1993a). Rafiei and Mendelzon (2000) develop similarity-based queries, using DFT. Janacek et al. (2005) propose to use likelihood ratio statistics to test the hypothesis of difference between series instead of an Euclidean distance in the transformed domain. Recent research uses wavelet transform to represent time series (Struzik and Siebes, 1998). In between, the discrete wavelet transform (DWT) has been found to be effective in replacing DFT (Chan and Fu, 1999) and the Haar transform is always selected (Struzik and Siebes, 1999; Wang and Wang, 2000). The Haar transform is a series of averaging and differencing operations on a time series (Chan and Fu, 1999). The average and difference between every two adjacent data points are computed. For example, given a time series $P=(1, 3, 7, 5)$, dimension of 4 data points is the full resolution (i.e. original time series); in dimension of two coefficients, the averages are (2 6) with the coefficients (-1 1) and in dimension of 1 coefficient, the average is 4 with coefficient (-2). A multi-level representation of the wavelet transform is proposed by Shahabi et al. (2000). Popivanov and Miller (2002) show that a large class of wavelet transformations can be used for time series representation. Dasha et al. (2007) compare different wavelet feature vectors. On the other hand, comparison between DFT and DWT can be found in Wu et al. (2000b) and Morchen (2003) and a combination use of Fourier and wavelet transforms are presented in Kawagoe and Ueda (2002). An ensemble-index, is proposed by Keogh et al. (2001b) and Vlachos et al. (2006), which ensembles two or more representations for indexing.

Principal component analysis (PCA) is a popular multivariate technique used for developing multivariate statistical process monitoring methods (Yang and Shahabi, 2005b; Yoon et al., 2005) and it is applied to analyze financial time series by Lesch et al. (1999). In most of the related works, PCA is used to eliminate the less significant components or sensors and reduce the data representation only to the most significant ones and to plot the data in two dimensions. The PCA model defines linear hyperplane, it can be considered as the multivariate extension of the PLR. PCA maps the multivariate data into a lower dimensional space, which is useful in the analysis and visualization of correlated high-dimensional data. Singular value decomposition (SVD) (Korn et al., 1997) is another transformation-based approach.

Other time series representation methods include modeling time series using hidden markov models (HMMs) (Azzouzi and Nabney, 1998) and a compression technique for multiple stream is proposed by Deligiannakis et al. (2004). It is based on base

signal, which encodes piecewise linear correlations among the collected data values. In addition, a recent biased dimension reduction technique is proposed by Zhao and Zhang (2006) and Zhao et al. (2006).

Moreover many of the representation schemes described above are incorporated with different indexing methods. A common approach is adopted to an existing multidimensional indexing structure (e.g. R-tree proposed by Guttman (1984)) for the representation. Agrawal et al. (1993a) propose an F-index, which adopts the R*-tree (Beckmann et al., 1990) to index the first few DFT coefficients. An ST-index is further proposed by (Faloutsos et al. (1994), which extends the previous work for subsequence handling. Agrawal et al. (1995a) adopt both the R*- and R+-tree (Sellis et al., 1987) as the indexing structures. A multi-level distance based index structure is proposed (Yang and Shahabi, 2005a), which for indexing time series represented by PCA. Vlachos et al. (2005a) propose a Multi-Metric (MM) tree, which is a hybrid indexing structure on Euclidean and periodic spaces. Minimum bounding rectangle (MBR) is also a common technique for time series indexing (Chu and Wong, 1999; Vlachos et al., 2003). An MBR is adopted in (Rafiei, 1999) which an MT-index is developed based on the Fourier transform and in (Kahveci and Singh, 2004) which a multi-resolution index is proposed based on the wavelet transform. Chen et al. (2007a) propose an indexing mechanism for PLR representation. On the other hand, Kim et al. (1996) propose an index structure called TIP-index (TIme series Pattern index) for manipulating time series pattern databases. The TIP-index is developed by improving the extended multidimensional dynamic index file (EMDF) (Kim et al., 1994). An iSAX (Shieh and Keogh, 2009) is proposed to index massive time series, which is developed based on an SAX. A multi-resolution indexing structure is proposed by Li et al. (2004), which can be adapted to different representations.

To sum up, for a given index structure, the efficiency of indexing depends only on the precision of the approximation in the reduced dimensionality space. However in choosing a dimensionality reduction technique, we cannot simply choose an arbitrary compression algorithm. It requires a technique that produces an indexable representation. For example, many time series can be efficiently compressed by delta encoding, but this representation does not lend itself to indexing. In contrast, SVD, DFT, DWT and PAA all lend themselves naturally to indexing, with each eigenwave, Fourier coefficient, wavelet coefficient or aggregate segment map onto one dimension of an index tree. Post-processing is then performed by computing the actual distance between sequences in the time domain and discarding any false matches.

3. Similarity measure

Similarity measure is of fundamental importance for a variety of time series analysis and data mining tasks. Most of the representation approaches discussed in Section 2 also propose the similarity measure method on the transformed representation scheme. In traditional databases, similarity measure is exact match based. However in time series data, which is characterized by its numerical and continuous nature, similarity measure is typically carried out in an approximate manner. Consider the stock time series, one may expect having queries like:

Query1: *find all stocks which behave "similar" to stock A.*

Query2: *find all "head and shoulders" patterns last for a month in the closing prices of all high-tech stocks.*

The query results are expected to provide useful information for different stock analysis activities. Queries like *Query2* in fact is tightly coupled with the patterns frequently used in technical analysis, e.g. double top/bottom, ascending triangle, flag and rounded top/bottom.

In time series domain, devising an appropriate similarity function is by no means trivial. There are essentially two ways the data that might be organized and processed (Agrawal et al., 1993a). In whole sequence matching, the whole length of all time series is considered during the similarity search. It requires comparing the query sequence to each candidate series by evaluating the distance function and keeping track of the sequence with the smallest distance. In subsequence matching, where a query sequence Q and a longer sequence P are given, the task is to find the subsequences in P, which matches Q. Subsequence matching requires that the query sequence Q be placed at every possible offset within the longer sequence P. With respect to *Query1* and *Query2* above, they can be considered as a whole sequence matching and a subsequence matching, respectively. Gavrilov et al. (2000) study the usefulness of different similarity measures for clustering similar stock time series.

3.1. Whole sequence matching

To measure the similarity/dissimilarity between two time series, the most popular approach is to evaluate the Euclidean distance on the transformed representation like the DFT coefficients (Agrawal et al., 1993a) and the DWT coefficients (Chan and Fu, 1999). Although most of these approaches guarantee that a lower bound of the Euclidean distance to the original data, Euclidean distance is not always being the suitable distance function in specified domains (Keogh, 1997a; Perng et al., 2000; Megalooikonomou et al., 2005). For example, stock time series has its own characteristics over other time series data (e.g. data from scientific areas like ECG), in which the salient points are important.

Besides Euclidean-based distance measures, other distance measures can easily be found in the literature. A constraint-based similarity query is proposed by Goldin and Kanellakis (1995), which extended the work of (Agrawal et al., 1993a). Das et al. (1997) apply computational geometry methods for similarity measure. Bozkaya et al. (1997) use a modified edit distance function for time series matching and retrieval. Chu et al. (1998) propose to measure the distance based on the slopes of the segments for handling amplitude and time scaling problems. A projection algorithm is proposed by Lam and Wong (1998). A pattern recognition method is proposed by Morrill (1998), which is based on the building blocks of the primitives of the time series. Ruspini and Zwir (1999) devote an automated identification of significant qualitative features of complex objects. They propose the process of discovery and representation of interesting relations between those features, the generation of structured indexes and textual annotations describing features and their relations. The discovery of knowledge by an analysis of collections of qualitative descriptions is then achieved. They focus on methods for the succinct description of interesting features lying in an effective frontier. Generalized clustering is used for extracting features, which interest domain experts. The generalized Markov models are adopted for waveform matching in Ge and Smyth (2000). A content-based query-by-example retrieval model called FALCON is proposed by Wu et al. (2000a), which incorporates a feedback mechanism.

Indeed, one of the most popular and field-tested similarity measures is called the "time warping" distance measure. Based on the dynamic time warping (DTW) technique, the proposed method in (Berndt and Clifford, 1994) predefines some patterns to serve as templates for the purpose of pattern detection. To align two time series, P and Q, using DTW, an n-by-m matrix M is first

constructed. The (ith, jth) element of the matrix, m_{ij}, contains the distance $d(q_i, p_j)$ between the two points q_i *and* p_j and an Euclidean distance is typically used, i.e. $d(q_i,p_j)=(q_i-p_j)^2$. It corresponds to the alignment between the points q_i and p_j. A warping path, W, is a contiguous set of matrix elements that defines a mapping between Q and P. Its kth element is defined as $w_k=(i_k, j_k)$ and

$$W = w_1, w_2, \ldots, w_k, \ldots, w_K \tag{2}$$

where $\max(m, n) \le K < m+n-1$.

The warping path is typically subjected to the following constraints. They are boundary conditions, continuity and monotonicity. Boundary conditions are $w_1=(1,1)$ and $w_K=(m,n)$. This requires the warping path to start and finish diagonally. Next constraint is continuity. Given $w_k=(a,b)$, then $w_{k-1}=(a',b')$, where $a-a' \le 1$ and $b-b' \le 1$. This restricts the allowable steps in the warping path being the adjacent cells, including the diagonally adjacent cell. Also, the constraints $a-a' \ge 0$ and $b-b' \ge 0$ force the points in W to be monotonically spaced in time.

There is an exponential number of warping paths satisfying the above conditions. However, only the path that minimizes the warping cost is of interest. This path can be efficiently found by using dynamic programming (Berndt and Clifford, 1996) to evaluate the following recurrence equation that defines the cumulative distance $\gamma(i,j)$ as the distance $d(i,j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements, i.e.

$$\gamma(i, j) = d(q_i, p_j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \tag{3}$$

A warping path, W, such that "distance" between them is minimized, can be calculated by a simple method

$$DTW(Q, P) = \min_W \left[\sum_{k=1}^{K} d(w_k)\right] \tag{4}$$

where $d(w_k)$ can be defined as

$$d(w_k) = d(q_{i_k}, p_{i_k}) = (q_{i_k}-p_{i_k})^2 \tag{5}$$

Detailed treatment can be found in Kruskall and Liberman (1983). As DTW is computationally expensive, different methods are proposed to speedup the DTW matching process. Different constraint (banding) methods, which control the subset of matrix that the warping path is allowed to visit, are reviewed in Ratanamahatana and Keogh (2004). Yi et al. (1998) introduce a technique for an approximate indexing of DTW that utilizes a FastMap technique, which filters the non-qualifying series. Kim et al. (2001) propose an indexing approach under DTW similarity measure. Keogh and Pazzani (2000b) introduce a modification of DTW, which integrates with PAA and operates on a higher level abstraction of the time series. An exact indexing approach, which is based on representing the time series by PAA for DTW similarity measure is further proposed by Keogh (2002). An iterative deepening dynamic time warping (IDDTW) is suggested by Chu et al. (2002), which is based on a probabilistic model of the approximate errors for all levels of approximation prior to the query process. Chan et al. (2003) propose a filtering process based on the Haar wavelet transformation from low resolution approximation of the real-time warping distance. Shou et al. (2005) use an APCA approximation to compute the lower bounds for DTW distance. They improve the global bound proposed by Kim et al. (2001), which can be used to index the segments and propose a multi-step query processing technique. A FastDTW is proposed by Salvador and Chan (2004). This method uses a multi-level approach that recursively projects a solution from a coarse resolution and refines the projected solution. Similarly, a fast DTW search method, an FTW is proposed by Sakurai et al. (2005) for efficiently pruning a significant number of search candidates. Ratanamahatana and Keogh (2005) clarified some points about DTW where are related to lower bound and speed. Euachongprasit and Ratanamahatana (2008) also focus on this problem. A sequentially indexed structure (SIS) is proposed by Ruengronghirunya et al. (2009) to balance the tradeoff between indexing efficiency and I/O cost during DTW similarity measure. A lower bounding function for group of time series, LBG, is adopted.

On the other hand, Keogh and Pazzani (2001) point out the potential problems of DTW that it can lead to unintuitive alignments, where a single point on one time series maps onto a large subsection of another time series. Also, DTW may fail to find obvious and natural alignments in two time series, because of a single feature (i.e. peak, valley, inflection point, plateau, etc.). One of the causes is due to the great difference between the lengths of the comparing series. Therefore, besides improving the performance of DTW, methods are also proposed to improve an accuracy of DTW. Keogh and Pazzani (2001) propose a modification of DTW that considers the higher level feature of shape for better alignment. Ratanamahatana and Keogh (2004) propose to learn arbitrary constraints on the warping path. Regression time warping (RTW) is proposed by Lei and Govindaraju (2004) to address the challenges of shifting, scaling, robustness and complexity. Latecki et al. (2005) propose a method called the minimal variance matching (MVM) for elastic matching. It determines a subsequence of the time series that best matches a query series by finding the cheapest path in a directed acyclic graph. A segment-wise time warping distance (STW) is proposed by Zhou and Wong (2005) for time scaling search. Fu et al. (2008a) propose a scaled and warped matching (SWM) approach for handling both DTW and uniform scaling simultaneously. Different customized DTW techniques are applied to the field of music research for query by humming (Zhu and Shasha, 2003; Arentz et al., 2005).

Focusing on similar problems as DTW, the Longest Common Subsequence (LCSS) model (Vlachos et al., 2002) is proposed. The LCSS is a variation of the edit distance and the basic idea is to match two sequences by allowing them to stretch, without rearranging the sequence of the elements, but allowing some elements to be unmatched. One of the important advantages of an LCSS over DTW is the consideration on the outliers. Chen et al. (2005a) further introduce a distance function based on an edit distance on real sequence (EDR), which is robust against the data imperfection. Morse and Patel (2007) propose a Fast Time Series Evaluation (FTSE) method which can be used to evaluate the threshold value of these kinds of techniques in a faster way.

Threshold-based distance functions are proposed by ABfalg et al. (2006). The proposed function considers intervals, during which the time series exceeds a certain threshold for comparing time series rather than using the exact time series values. A T-Time application is developed (ABfalg et al., 2008) to demonstrate the usage of it. Fu et al. (2007) further suggest to introduce rules to govern the pattern matching process, if a priori knowledge exists in the given domain.

A parameter-light distance measure method based on Kolmogorov complexity theory is suggested in Keogh et al. (2007b). Compression-based dissimilarity measure (CDM)[3] is adopted in this paper. Chen et al. (2005b) present a histogram-based representation for similarity measure. Similarly, a histogram-based similarity measure, bag-of-patterns (BOP) is proposed by Lin and Li (2009). The frequency of occurrences of each pattern in

[3] CDM is proposed by Keogh et al. (2004), which is used to compare the co-compressibility between data sets.

the time series is counted and compared by CDM. Lang et al. (2010) develop a dictionary compression score for similarity measure. A dictionary-based compression technique is suggested to compute long time series similarity.

3.2. Subsequence matching

In subsequence matching where a query sequence and a longer time series are given, the task is to find the subsequences in the longer time series, which matches the query sequence. The query sequence is required to place at every offset within the longer time series. Faloutsos et al. (1994) generalizes the work in Agrawal et al. (1993a) for subsequence searching. Based on this work, many researches are conducted to improve the performance of the subsequence searching. For example, a DualMatch is proposed by Moon et al. (2001) to divide the time series into disjoint windows and query pattern into sliding windows. Loh and Kim (2001) extend (Faloutsos et al., 1994) using an index interpolation to solve the storage and CPU time overhead.

The GeneralMatch method is proposed by Moon et al. (2002), which reduces the window size effect by using large windows by the method in Faloutsos et al. (1994) and exploits point-filtering effect by DualMatch (Moon et al., 2001). Furthermore, Kim and Jeong (2007) discuss on the potential performance bottleneck during subsequence matching. Four major areas, that time required to process subsequence matching, are identified. They are processing time, disk access time and the corresponding post-processing steps of them. A window ordering method is proposed to eliminate the redundancies of disk access and CPU processing in the post-processing steps. A method based on an index interpolation is further proposed by Lim et al. (2007) to improve the performance of DualMatch.

Moreover subsequence matching using sequence of linear segments (Morinaka et al., 2001), anomaly subsequence detection using an SAX (Keogh et al., 2005), online subsequence matching using PLR (Wu et al., 2004) and weighted subsequence matching using PLR (Wu et al., 2005) can be found in the literature. All these approaches including the DFT approaches are based on lower bounding of the Euclidean distance.

Li et al. (1996) propose a hierarchical similarity search algorithm for locating subsequences in the transformed domain (e.g. DFT, wavelet) hierarchically. An indexing method, S^2-Tree, is presented by Wang and Perng (2001) for subsequence matching, which is based on string searching techniques on different time series representation schemes.

In the context of subsequence searching by DTW, a suffix tree (Gusfield, 1997) is proposed to index the DTW for subsequence matching (Park et al., 1999; Park et al., 2000; Kim et al., 2002). Other approaches include a segment-based approach based on piecewise time warping (Park et al., 2001a), an index-based approach based on prefix querying (Park et al., 2001b) and an optimization approach (Kim et al., 2005).

Han et al. (2007) develop a ranked subsequence matching algorithm to reduce the number of subsequence needs to access by defining the minimum-distance matching-window pair (MDMWP). Directed acyclic graph (DAG) is adopted in Dorr and Denton (2009) to capture the relationship between subsequences and patterns.

Other researches which are related to time series similarity measure and pattern matching include forecasting by pattern recognition (Singh and Stuart, 1998; Liu et al., 2004) and defining query language. Shape definition query (SDL) is introduced by Agrawal et al. (1995b) for retrieving objects based on the shapes contained in the histories associated with these objects. Jagadish et al. (1995) propose a framework, which include a pattern language, a transformation rule language and a query language, for defining queries in terms of similarity of objects. Lin and Risch (1998) extend the SELECT operator in SQL that retrieves implicit values from a discrete time sequence under various user-defined interpolation assumptions. Anand et al. (2001) propose a chart-pattern language (CPL) to enable financial analysts to define patterns with subjective criteria and incrementally compose complex patterns from simpler patterns for pattern query. Dong et al. (2009) propose to measure the shape distance of the time series. The shapes are described according to the relative changes of the slopes lines. Finally, detail comparisons or experiments on the existing time series representation and similarity measure approaches can be found in Keogh and Kasetty (2002) and Ding et al. (2008).

4. Segmentation

Time series segmentation can be considered either as a preprocessing step for variety of data mining tasks or as trend analysis techniques. It is also considered as a discretization problem. Unlike transactional databases with discrete items, time series data is characterized by their numerical and continuous nature. In Das et al. (1998), a simple discretization method is proposed. A fixed length window is used to segment a time series into subsequences and the time series is then represented by the primitive shape patterns that are formed. This discretization process mainly depends on the choice of the window width. However, using fixed-length segmentation is an over-simplified approach to solve the problem. There are at least two identified disadvantages. First, meaningful patterns typically appear with different lengths throughout a time series. Second, as a result of the even segmentation of a time series, meaningful patterns may be missed if they are split across time (cutting) points. Thus, it is better to use a dynamic approach, which identifies the time points in a more flexible way (i.e. using different window widths).

This is certainly not a trivial segmentation problem. Common segmentation methods include using the PIP (Fu et al., 2006; Jiang et al., 2007) or detecting special events (Guralnik and Srivastava, 1999) in the time series as the time points, minimum message length (MML) (Oliver et al., 1998) and minimum description length (MDL) segmentation (Fitzgibbon et al., 2002). Fancourt and Principe (1997) adopt PCA for the segmentation problem. Based on PCA, a fuzzy clustering based segmentation is proposed by Abonyi et al., (2003, 2005). A two stages approach which first uses piecewise generalized likelihood ratio (GLR) to rough segmentation and then refines the results is proposed by Wang and Willett (2004). On the other hand, Keogh et al. (2001c) adopt PLR to segment the time series. They focus on the problem of an online segmentation of time series and a sliding window and bottom-up (SWAB) approach is proposed.

Oliver and Forbes (1997) suggest that the time points are identified at which behavior changes occur in a time series. In the statistical term, this is called the "change-point detection problem". The standard solution involves fixing the number of change-point, then identifying their positions, and finally determining functions for curve fitting the intervals between successive change-points. Chu (1995) presents a sliding test window segmentation procedure which is based on non-stationary detection on fluctuation statistics and change-point localization. An iterative algorithm is proposed by Guralnik and Srivastava (1999) that fits a model to a time segment and then uses a likelihood criterion to determine if the segment should be partitioned further. Srivastava and Weigend (1996) suggest discovering the underlying switching process in a time series, which entails identifying the number of sub-process and the

dynamics of each sub-process. The concept of the nonlinear gated experts derived from statistical physics is proposed to perform the segmentation. In Duncan and Bryant (1996), dynamic programming is proposed to determine the total number of intervals within the data, the location of these intervals and the order of the model within each segment. In (Srivastava et al., 1999), the segmentation problem is considered with a tool for exploratory data analysis and data mining called the scale-sensitive gated experts (SSGE), which can partition a complex nonlinear regression surface into a set of simpler surfaces called "features". An improved annealed competition of experts algorithm (ACE) identifies switching dynamics in time series using on mutual information and false nearest neighbor to determine appropriate embedding dimension and time delay (Feng et al., 2005).

The segmentation problem has also been considered from the perspective of finding cyclic periodicity for all of the segments. In Han et al. (1998, 1999), the data cube and the Apriori data mining techniques are used to mine segment-wise periodicity, using a fixed length period. An off-line technique for the competitive identification of piecewise stationary time series is described by Fancourt and Principe (1996). In addition to performing piecewise segmentation and identification, the proposed technique maps similar segments of a time series as neighbors on a neighborhood map.

Himberg et al. (2001) propose a global iterative replacement (GIR) method, which approximates the dynamic programming result for minimizing the intra segment variances. The proposed method is applied to context recognition for the mobile phone applications.

Although the approaches described in this section can generally identify a given pattern from a time series, they do not consider the problem of identifying a suitable set of time points in a time series, when a set of pattern templates is given; for example, the technical patterns (e.g. H&S, double top, etc.) for the stock analysis. Further, in order to form a versatile mining space, a variety of patterns (e.g. in different resolutions) have to be identified. The aforementioned segmentation task can be regarded as an optimization problem and Chung et al. (2004) propose a solution, which is based on an evolutionary computation.

5. Visualization

Visualization is an important mechanism to present the processed time series for further analysis by users. It is also a powerful tool to facilitate the mining tasks like pattern searching, query-by-example, and pattern discovery afterwards. Current tools for visualizing time series include: (1) cluster and calendar-based visualization tool (van Wijk and van Selow, 1999), which obtains chunks of data with a given interval and then clusters them accordingly and (2) spiral visualization tool (Weber et al., 2001), which maps the periodic section of time series into a ring. These two tools are focused on periodic time series and a fixed length of period must be provided, say, weekly or monthly. A financial visual analytics system for pattern-based analysis of 2-dimensional time-vary chart data is proposed by Schreck et al. (2007). Hao et al. (2007) introduce the notion of degree of interest (DOI) to define and generate multi-resolution layouts of long time series. Non-linear rescaling and space-efficient rendering method are used to visualize the long time series.

Keogh et al. (2002a) and Hochheiser and Shneiderman (2004) developed a tool called TimeSearcher which is a time series exploratory and visualization tool, so that a user can retrieve time series by querying. Based on their previous developed TimeBoxes, which are rectangular, direct-manipulation time series queries, they extend it by introducing variable time timeboxes (VTT), which permits the specification of queries to allow uncertainty in the time axis. Four methods (sample events, aggregated sample events, event index, and interleaved event index) to represent the unevenly space time series data are studied in Aris et al. (2005). Furthermore, TimeSearcher2 is developed by Buono et al. (2005), in which a new search interface combining both filter and pattern search capability is provided. TimeSearcher is focused on multiple time series query based on examples. Specification of the region of interest must be provided.

Recently, another time series visualization tool called VizTree is proposed (Lin et al., 2005a). This approach first converts each numeric time series to a symbol string based on the SAX and a set of substrings (with the same number of symbol) extracted from the symbol string is encoded by a modified suffix tree to visualize the frequency of patterns. That is, the SAX discretizes the original time series into fixed length subsequences, converts each subsequence to a symbol and the symbols obtained are concatenated to form a symbol string. Given a symbol string, say *abccbccbcabcbcc*, the next step is to convert this long string to a set of substrings according to the length of each substring, W (or the number of substring), specified by users. For example, if the preferred length of substring is 3, the given string will be divided into 5 substrings if jumping window is used, i.e. *abc*, *cbc*, *cbc*, *abc* and *bcc*. Similarly, 13 substrings will be obtained if the sliding window is used, i.e. *abc*, *bcc*, *ccb*, *cbc*, *bcc*, *ccb*, *cbc*, *bca*, *cab*, *abc*, *bcb*, *cbc* and *bcc*. Next, a suffix tree will be constructed as exemplified in Fig. 4. The length of substring is reflected by the depth of the tree. Each branch of the tree represents a pattern. The frequency of the pattern is represented by the thickness of each branch. As shown in Fig. 4, frequently appearing patterns discovered by using jumping window are *abc* and *cbc* (Fig. 4a), while *bcc* and *cbc* are discovered by using the sliding window (Fig. 4b). Different applications of VizTree are suggested by the authors including subsequence matching, frequently appearing pattern discovery and surprising pattern discovery.

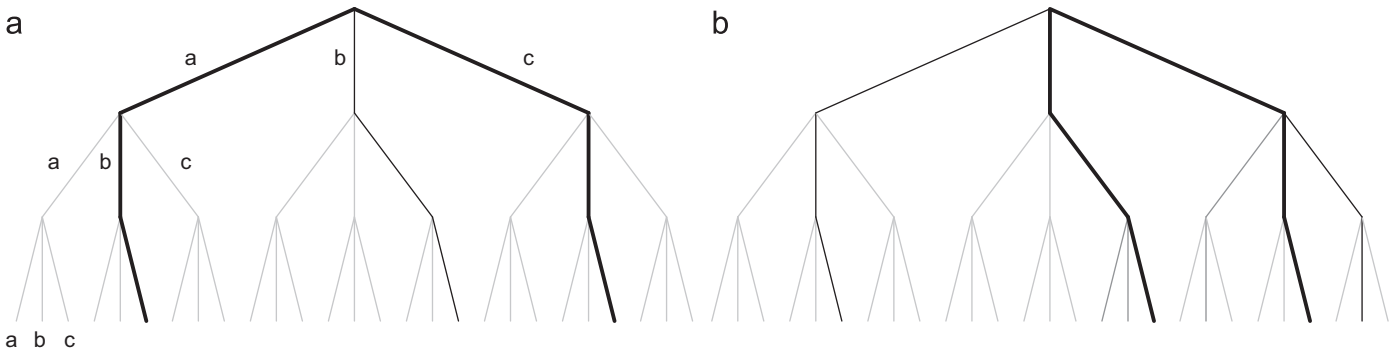

Fig. 4. VizTree constructed using (a) jumping window and (b) sliding window.

The authors demonstrated that VizTree is capable of discovering frequently appearing and surprising patterns in a given resolution, which is suitable for time series applications like an ECG. The same representation (i.e. SAX) but different visualization tools are also proposed by using bitmap (Kumar et al., 2005) and dot plots (Yankov et al., 2005). On the other hand, Fu et al. (2008b) extend the work to the discovery of interesting patterns across different resolutions by adopting the symbolic representation of PIP instead of an SAX in the VizTree.

6. Mining in time series

Mining is the final goal to discover hidden information or knowledge from either the original or the transformed time series data. Indeed, pattern discovery is the most common mining task and the clustering method is the most commonly method. Other time series data mining tasks include classification, rule mining and summarization.

6.1. Pattern discovery and clustering

It is a non-trivial task to discover interesting patterns, which include frequently appearing (Fu et al., 2001) and surprising patterns (Keogh et al., 2002b), from time series data. These tasks are also called motif discovery (Chiu et al., 2003; Tanaka et al., 2005) and anomaly detection (Chan and Mahoney, 2005; Wei et al., 2005) or finding discords (Keogh et al., 2007a), respectively. The discovery of interesting patterns has become one of the most important data mining tasks, and it can be applied to many domains (Caraa-Valente and Lopez-Chavarrias, 2000; Lerner et al., 2004). Ma and Perkins (2003) present a support vector regression (SVR)-based online novelty detection algorithm. Chan and Mahoney (2005) present an online anomaly detection approach based on the Gecko algorithm, which creates a sequence of minimal bounding boxes with the training trajectories.

For the problem of time series pattern discovery, a common group of techniques being employed is distance-based clustering (Das et al., 1998; Oates, 1999; Wang et al., 2002). The general clustering procedure is listed in Fig. 5. In each iteration, the winner cluster is found and its center is updated accordingly. The initial cluster centers can be chosen in various ways, e.g. chosen arbitrarily or by some sequences. Also, the number of cluster is a critical parameter to be determined. It can be fixed beforehand or can vary during the clustering process. The clustering procedure finally terminates when the number of iteration exceeds the maximum allowed number of iterations or convergence.

While patterns can be directly discovered from time series, a major problem is that time series data mostly increase linearly with time. This will cause the storage needs to increase rapidly and slow down the pattern discovery process exponentially. Therefore, an effective mechanism for compressing the huge amount of time series data, especially historical data, is needed. This not only reduces the size of storage, but also maintains an acceptable level of information for the discovery process. Fu et al. (2001) propose to adopt PIP representation to solve these problems. A neural clustering method, the self-organizing map (SOM) (Kohonen, 1995), is used for pattern discovery. An SOM is a special type of clustering algorithm (Ripley, 1996) with an immense discovery power. Spatio-temporal self-organizing feature maps are proposed by Euliano and Principe (1996). Other researches adopt an SOM for time series data including: Ultsch (1999) and Morchen et al. (2005) adopt emergent feature maps in the medical domain, Kuo et al. (2004) focus on financial domain and use the K-chart (i.e. combining open, high, low and close prices) analysis as the input of the SOM for prediction and Wang et al. (2005b) propose a dimensionality reduction method using global characteristics like trend and seasonality, periodicity, skew for feature selection before using an SOM. Guo et al. (2007) adopt an SOM for stock pattern discovery. An improved version of the rival penalized competitive learning (RPCL) is further introduced.

Moller-Levet et al. (2003) adopt the fuzzy *c*-means (FCM) algorithm for short time series and unevenly spaced sampling points' time series clustering. They propose to measure the similarity of short time series based on shapes, which are formed by the relative change of amplitude and the temporal information. Bargiela and Pedrycz (2003) discuss the notion of granular data, elaborate on the recursive information granulation and access the quality of summarization of information granules through an FCM clustering. Steinbach et al. (2003) present a clustering-based method to discover climate indices that represent regions with relatively homogeneous behavior. Lin et al. (2004) propose an anytime version of partitioned clustering algorithm, which adopts the multi-resolution property of wavelets. Anytime algorithm means trade execution time for quality of results and always has a best-so-far answer available and the quality of the answer improves with an execution time. The user may examine this answer at anytime (Grass and Ziberstein, 1996). Similarly, Lin et al. (2005b) introduce another multi-resolution clustering approach based on multi-resolution PAA (MPAA) for the iterative clustering algorithm of streaming time series.

Autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models have also been used extensively for time series analysis. Kalpakis et al. (2001) propose to cluster ARIMA time series, using the partition around the medoids method. Xiong and Yeung (2004) focus on the problem of clustering time series of different lengths, using mixtures of ARMA models and expectation-maximization (EM) algorithm.

```
Function clustering
Begin
   Define number of cluster centers
   Set the initial cluster centers
   Repeat
      For each input data
         For each cluster, compute:
             Distance between the input data and its center
             Choose the closest cluster center as the winner
             Update the cluster centers
         End For
      End For
   Until (Convergence or maximum number of iterations is reached)
End
```

Fig. 5. A typical clustering process.

Bagnall and Janacek (2005) focus on clustering data derived from ARMA models, using *k*-means and *k*-medoids algorithms. A clipping process, which descretizes data into binary sequences of above and below the media, is adopted. This process is strengthens on the presence of an outlier in the data.

Hidden Markov model (HMM) is a common model-based algorithm adopted in time series clustering (Panuccio et al., 2002). HMMs are defined as stochastic generalizations of finite-state automata, where both transitions between states and generation of output symbols are governed by probability distributions. Oates et al. (1999) present a hybrid time series clustering algorithm that uses DTW for rough initialization and HMM for removing the sequences that do not belong to the clusters. Yin and Yang (2005) propose to transform the sensor time series data into an equal-length vector and model it as a HMM for spectral clustering. Furthermore, a recursive HMM training process is proposed by Duan et al. (2005).

Finding representative trends is a typical task, which belongs to the time series pattern discovery area. Indyk et al. (2000) identify various representative trends in time series over arbitrary windows of interest. An approximation approach based on replacing each interval by a "sketch", which is a low dimensional vector, is adopted. Papadimitrious and Yu (2006) also examine the time series at multiple time scales and discover the key trend in each (i.e. the optimal local patterns). Udechukwu et al. (2004) propose to convert the time series to symbolic form, build the suffix tree and discover the frequent patterns or trends accordingly.

On the other hand, discovering periodic patterns is another common focus for pattern discovery. Han et al. (1998, 1999) propose methods to discover periodicity or partial periodicity segments. Berberidis et al. (2002a) search for weak periodic signals using autocorrelation function and fast Fourier transform (FFT) with no period length are known in advance. This work extends the work from the same authors in (Berberidis et al., 2002b). Elfeky et al. (2004) further develop a one-pass algorithm based on convolution. A summary of the work by these authors can be found in Elfeky et al. (2005). Vlachos et al. (2004) apply power spectral density estimation using DFT and tree index to discover important periods. It is applied on an identification of bursts for online search queries.

Cluster analysis is also applied to the sliding window of the time series for grouping related time series subsequence patterns that are dispersed along the time series. Clustering methods seek out a special type of structure, namely, grouping tendencies in the data. In this regard, they are not as general as the other approaches, but they can provide valuable information when local aggregation of the data is suspected. Das et al. (1998) propose that the pattern templates for matching are not predefined. Instead, the templates are generated automatically by clustering techniques and they will then be used for further matching in the discretization process to produce meaningful symbols. Policker and Geva (2000) describe adaptive methods for finding rules of the above type from time series data. Methods are based on discretizing the sequence by methods resembling vector quantization. Again, they first form subsequences by sliding a window through the time series, and then cluster these subsequences by using a suitable measure of time series similarity. Denton (2005) proposes a kernel-density-based clustering for time series subsequences.

However, applying clustering approaches to discover frequently appearing patterns is claimed to be meaningless when focusing on time series subsequence recently (Keogh et al., 2003). It is because when using a sliding window to discretize the long time series into subsequences in a fixed window size, patterns, which are derivations from sine curve, are always resulted no matter how the shape of the given time series is. Theoretical analysis and experiments can be found in Wang et al. (2005a). Lin et al. (2002) state that the definition of a match is rather obvious and intuitive; but it is needed for the definition of a trivial match as it is easy to observe that the best matches to a subsequence tend to be the subsequence that begin just one or two points to the left or the right of the subsequence. They define the term trivial match as: given a time series P containing a subsequence S_1 beginning at position p and a matching subsequence S_2 beginning at q, S_2 is a trivial match to S_1 if either $p=q$ or there does not exist a subsequence S_2' beginning at q' such that $D(S_1, S_2') > R$, and either $q < q' < p$ or $p < q' < q$. Therefore, it is necessary to prevent the over-counting of these trivial matches.

Lin et al. (2002) and Patel et al. (2002) define the problem of enumerating the most frequently appearing patterns in a time series P (Lin et al. (2002) referred to as the most significant motifs, 1-motif) is the subsequence S_1 that has the highest count of non-trivial matches. Therefore, the Kth most frequently appearing patterns (significant motif, K-Motif) in P is the subsequence S_K that has the highest count of non-trivial matches and satisfies $D(S_K, Si) > 2R$, for all $1 \le i < K$.

In addition, Chiu et al. (2003) address the limitations discussed in Lin et al. (2002) on poor scalability of the motif discovery algorithm and the inability to discover motifs in the presence of noise by applying a probabilistic model to the algorithm. However, it is still difficult to define a threshold, R, to distinguish trivial and non-trivial matches as it is case dependent and there is no general rule for defining this value.

Furthermore, Keogh et al. (2003) suggest applying a classic clustering algorithm to cluster only the motifs discovered from K-motif detection algorithm in place of subsequence time series clustering. It is because a subset of the motifs discovered might really be a group that should be clustered together to extract promising subsequences from the data. The motif-based-clustering algorithm is as shown in Fig. 6.

Chen (2005) shows that time series clustering is not meaningless when using delay space method instead of an Euclidean distance for the similarity measure. Fu et al. (2005a) propose an intermediate subsequences filtering process by detecting the change of PIP before the clustering process. Simon et al. (2006) introduce an unfolding (subsampling) preprocess method before the subsequence clustering, using an SOM. Furthermore, another solution is proposed by Chen (2007), which is based on restricting the clustering space. A disk-aware algorithm is proposed by

```
Function motif-based-clustering
Begin
  Decide on a value for k
  Discover the K-motifs in the data, K>>k, i.e. K=k×c
  (c is some constant, in the region of about 2 to 30)
  Run k-means, or k hierarchical clustering, or any other clustering
  algorithm on the subsequence covered by K-motifs
End
```

Fig. 6. Pseudo code of the motif-based-clustering algorithm.

Mueen et al. (2009) to find exact motifs in massive time series databases.

Interesting pattern discovery also includes the detection of anomalies. Vlachos et al. (2005c) focus on non-parametric methods that extract important periodic features for classification and anomaly detection and Keogh et al. (2006) adopt the SAX representation to improve the performance of finding surprising patterns. Wei et al. (2006) adopt the SAX representation for anomaly detection on the basis of shape. Based on an SAX, suffix tree and Markov model, surprising patterns are discovered in Lonardi et al. (2006) if the frequency of their occurrence differs substantially from the expected chance of existing patterns. On the other hand, a bitmap approach is also proposed by Wei et al. (2005). Yankov et al. (2007) focus on discovering surprising patterns under uniform scaling. Yankov et al. (2008) further focus on finding surprising patterns in terabyte sized data sets. Two linear scans through the database are adopted to reduce the memory usage.

Most of the approaches, discussed so far, are focused on discovering patterns in a fixed resolution (i.e. fixed period), Bettini et al. (1998) and Li et al. (2000b) focus on discovering patterns across different resolutions, which are called multiple granularities. Bettini et al. (1998) propose timed automata with granularities (TAGs), while calendar schemata is presented by Li et al. (2000b). A *k*-motif-based algorithm is proposed by Tang and Liao (2008) for discovering patterns with different lengths.

6.2. Classification

Classification is a traditional data mining task. In the time series domain, special treatment must be considered due to the nature of the data. Geurts (2001) proposes to classify time series data based on combining local properties or patterns in the time series. Zhang et al. (2004) develop a representation method using wavelet decomposition that can automatically select the parameters for the classification task. They propose a nearest neighbor classification algorithm, using the derived appropriate scale. Kadous and Sammut (2005) use metafeature approach (i.e. recurring substructure) like local maxima in time series to generate classifiers. Similarly, Yang et al. (2005) focus on feature subset selection (FSS) based on common principal components, which is called CleVer, to retain the correlation information among original features. Classification is employed to evaluate the effectiveness of the selected subset of features.

On the other hand, researchers have also focused on customizing or developing classifiers for time series data. For example, Povinelli et al. (2004) present a signal classification approach based on modeling the dynamics of a system as they are captured in a reconstructed phase using Gaussian Mixture models of time domain signatures. Rodriguez and Alonso (2004) study both interval and DTW-based decision trees adequate for the classification of time series data. Ensembles are used to combine base classifiers, while Wei and Keogh (2006) study the combination of the numerosity reduction, using DTW and nearest-neighbor classifiers for time series classification. Also, Xi et al. (2006) propose a semi-supervised time series classifiers when only a small set of labeled examples is available.

6.3. Rule discovery

Rule mining is another typical task in the field of data mining. Association rule mining (Agrawal et al., 1993b; Agrawal and Srikant, 1994) is one of the most well known algorithms. However, it is mainly focused on symbolic items present in transactions. Therefore, many researchers propose new or modified algorithms for rule mining in the context of time series data.

A trivial approach is first discretizing the time series data into segments and converting each segment to a symbol. Then rules can be discovered in the transformed symbolic domain. Das et al. (1998) cluster the subsequences to find the symbols, and then apply simple rule mining method to discover the hidden rules. Lu et al. (1998) propose *n*-dimensional inter-transaction association rules for handling spatial and multimedia data mining. Last et al. (2001) focus on discovering fuzzy association rules, which is based on the computational theory of perception and signal processing techniques. Leigh et al. (2002) develop a financial rule discovery method, using "bull flag" technical charting heuristics. Similarly, Ting et al. (2006) propose to representation financial time series based on candle stick charting technique for rule mining.

The research group of Hoppner (Hoppner and Klawonn, 2001) takes durations into account and develops a framework to discover temporal rules, which have been generated out of a set of frequent patterns in a state sequence. The framework represents the segments of time series by attributes (e.g. increasing, high-value, highly convex) and discovers interval relationships described in terms of an interval logic. Another research group of Hetland and Saetrom (2005) present a rule mining method that is based on genetic programming and specialized hardware. They also examine the role of discretization when evolving time series predictor rules.

Besides using association rules, decision tree is another common approach for rule mining. Ohsaki et al. (2003) first discuss on the preprocessing step to discover interesting rules from the medical time series data. Similar to Das et al. (1998), clustering is adopted to discover typical patterns from subsequences for the discretization process. The rule mining method is based on pattern extraction and decision tree. Cotofrei and Stoffel (2002) propose formalism based on the temporal first logic-order for rule mining. The approach first transforms sequential raw data into sequences of events, then infers temporal rules using the classification trees.

In addition, Jin et al. (2002) focus on discovering the distribution of temporal rules and Sarker et al. (2003) focus on developing parallel algorithm for time series rule mining.

6.4. Summarization

Some of the researches are focused on summarizing and describing the time series data for analysis, mining or prediction. Zwir and Enrique (1999) develop an automated identification of significant qualitative features (interesting patterns) in complex objects (time series). Clustering techniques are adopted to summarize and produce a compact description of salient features and their relations.

Boyd (1998) developed a system that integrates knowledge-based signal processing and natural language processing to automatically generate descriptions, and it is tested on the weather data. These descriptions are based on short and long-term trends, which are detected using the wavelet transform. Guimaraes and Ultsch (1999) propose an approach to transit patterns in multivariate time series to a linguistic description. Different abstraction levels' temporal grammatical rules are extracted from the results of neural networks and other exploratory methods. This approach is applied to medical time series, i.e. sleep-related breathing disorders (Guimaraes et al., 2001).

The SumTime project is carried out by Sripada et al. (2001, 2003), which aims to develop generic techniques for summarizing

time series data. The developed system can generate English textual summaries of time series data by selecting the most important trends and patterns, mapping these patterns onto words and phrases and generating actual texts based on these words and phrases. SumTime-Turbine is developed by Yu et al. (2004), which is focused on the sensor data from gas turbines. Reiter et al. (2005) developed a SumTime-Mousam weather-forecast generator, which uses consistent data-to-word rules.

Ahmad et al. (2004b) present a time series summarization method based on analyzing non-stationary, volatile and high-frequency time series data. Multi scale wavelet analysis is used to separate the trend, cyclical fluctuations and autocorrelation effects. It helps to provide a summary of the data with respect to the "chief features" of the data. The framework can generate text signals to describe each effect. Ahmad et al. (2004a) outline a system that comprises modules for summarizing texts and time series to study the link between them. Similarly, previous work by Lavrenko et al. (2000) demonstrate how to use language models to associate stories and trends in time series. They identify trends in time series using PLR and use language models to represent patterns of language that are highly associated with particular trends. All these researches are evaluated using data from financial domain (e.g. stock market, currency rate). A visualization system is developed and described with a financial trading case study by Taskaya and Ahmad (2003).

Furthermore, Kacprzyk et al. (2008) propose to use the fuzzy quantifier to present a linguistic summarization on the trends of time series which the trends are identified by PLR. Similarly, Batyrshin and Sheremetov (2008) describe a perception-based decision making system which time series is represented by fuzzy sets of perceptions. A linguistic scaling of patterns is used to define the vocabulary.

6.5. Recent research directions

We discussed four major time series data mining tasks so far; they are: pattern discovery (clustering), classification, rule discovery and summarization. Due to the mature development in this field and the significant enhancement on the hardware and communication technologies, three extensions attract more researchers focused on recently. They are mining on multi-attribute time series, mining on time series data stream and also the privacy issue. Some researches discussed above also proposed partial solutions or directions on them.

First, multi-attribute time series data can also be considered as multiple time series manipulation. Povinelli and Feng (1999) propose an approach which temporal clusters from multiple time series are used and a genetic algorithm is adopted. The method reconstructs state space for temporal pattern extraction and adopts an optimal local model for short-term forecasting. The performance of the approach is demonstrated by using financial non-stationary time series (i.e. stock price and volume). Kahveci et al. (2002) consider the problem of shift and scale invariant search for multi-attribute time series. A symmetric distance function and a Cone Slice (CS) index are proposed. Lee et al. (2009) focus on mining of closed patterns in multi-sequence time series by adopting an SAX representation.

Recent researches also focus on mining multivariate time series data. Minnen et al. (2007a, 2007b) propose different algorithms to discover multivariate frequently appearing patterns (i.e. motif discovery). Tatavarty et al. (2007) consider the problem of discovering the temporal dependencies between the frequently appearing patterns in multivariate time series. Wang et al. (2007) and Plant et al. (2009) focus on the clustering issue, while Takashi et al. (2009) focus on the prediction issue.

Second, data streaming is referring to transfer of data at a steady high-speed rate. Handling corresponding time series data received considerable attention recently, because of the increase in network bandwidth and its stability. It differs from traditional time series data on its characteristics of huge amount of data arriving at steady high-speed rate. One of the initial works on time series data streaming mining mainly focuses on the design of system architecture. Miller et al., (1998) propose an *I*/*O* system design and implementation targeted at applications, which perform data streaming. Golab and Ozsu (2003) reviewed recent work in data stream management systems with an emphasis on application requirements, data models, continuous query languages and query evaluation. Chen et al. (2002) investigate methods for an online multi-dimensional regression analysis of time series stream data. They show that only a small number of compressed regressions need to be measured instead of a complete stream of data for a multi-dimensional linear regression analysis. Lian and Chen (2008) propose a framework to handle similarity search, values prediction and indexing over data stream.

Based on the well-developed time series data mining algorithms in different aspects, they are either applied directly or customized for streaming time series data. Indeed, representation of streaming time series for dimensionality reduction and an online query or matching is a hot topic. It is important that an incoming stream of data is a continually appended time series in a database. Each time when a new data point arrives, the system needs to fetch/get the data from the database, the nearest or the neighboring data of the incoming time series is up to the time position and most researches focus on investigating the correlation of the data. Yi et al. (2000) develop a fast method to analyze the co-evolving time series for estimating and forecasting, quantitative data mining and outlier detection. Gilbert et al. (2001) adopt sketch based methods for capturing various linear projections of the data for representing data streams (i.e. wavelet transform) and approximate aggregate query. Gao and Wang (2002) tackle the problem by using an FFT to find the cross correlations of time series in a batch mode efficiently. Gao et al. (2002) focus on continuous nearest neighbor query, using existing indexing methods with pre-fetching. Cole et al. (2005) propose to combine several simple techniques (e.g. sketches, convolution, structured random vectors, etc.) to compute Pearson correlation over uncooperative time series. Vlachos et al. (2005b) examine the problem of monitoring and identifying correlation burst patterns in multi-steam time series data. The solution adopts burst detection and indexing. Ogras and Ferhatosmanoglu (2006) developed a transformation-based framework to reduce the dimension for large-scale and dynamic time series data online. The framework is focused on DFT-based synopsis generation and a recursive method is introduced to update the highest energy transform coefficients of the series data. Wei et al. (2007) introduce an on-the-fly subsequence matching of streaming time series to a set of predefined patterns, using the filtering approach. The proposed approach merges similar patterns into a wedge, which is an envelope-based lower bounding technique, to speedup the matching process. A stream-DTW (STW) distance measure is proposed by Capitani and Ciaccia (2007) for continuously monitoring DTW distance measure of time series data streams. Boolean representation based on the data-adaptive correlation analysis is proposed by Zhang et al. (2007). Palpanas et al. (2008) introduce arbitrary user-specified amnesic functions based on PLR to allow an online approximation of streaming time series. This function allows arbitrary, user-defined reduction of quality with time. A tree structure is further proposed by Fu et al. (2008c) for storing the PIPs, which supports various incremental updating approaches (Fu et al., 2005b). A multiscale segment

mean (MSM) approximation is proposed by Lian et al. (2009) which support incrementally computation and static/dynamic pattern matching.

A tutorial presenting techniques for finding sliding window correlations, discovering bursts, matching hums and maintaining and manipulating time ordered data stream can be found in Lerner et al. (2004). Chen et al. (2007b) propose to handle continuous pattern detection, using spatial assembling distance (SpADe). An SpADe is proposed in this paper to handle both shifting and scaling in temporal and amplitude dimensions. Lim et al. (2008) concentrate on continuous query sequences on time series data stream based on window construction mechanism for supporting variable length queries. Two online segmentation methods, (stepwise) feasible space window (FSW/SFSW), are proposed by Liu et al. (2008) to improve the performance of classic sliding window method. A distortion-free predictive streaming time series matching algorithm is introduced by Loh et al. (2010). The proposed algorithm performs preprocessing step to remove distortions and predict future search results simultaneously.

Furthermore, researchers also extend their interest on mining time series streaming data. Yamanishi and Takeuchi (2002) present an online learning framework based on a probabilistic model for outlier detection and change-point detection on the time series data stream. Papadimitriou et al. (2005) introduce a streaming pattern discovery method in multiple time series, which summarizes the key trends in the stream collection based on PCA. An online clustering system is proposed by Rodrigues et al. (2008) which a top-down strategy is adopted to construct a binary tree hierarchy of clusters. Clusters' diameters are evolved continuously with the stream data.

Third, research on data mining is suggested to incorporate with privacy concern (Agrawal and Srikant, 2000). Working on the privacy in time series data mining is a newly research direction. Zhu et al. (2008) suggest that traditional techniques are not effective in the time series data. Data flow separation attack is identified and possible countermeasures to this attack are further proposed in this paper. To preserve the privacy, cloaked time series data are suggested to be adopted. Lian et al. (2008) propose an approach to deal with cloaked range query (CRQ) based on an R-tree indexing. Furthermore, framework is proposed by Nin and Torra (2009) to evaluate different time series protection methods. A set of information loss and disclosure risk measures for time series are introduced in this paper. Based on the definition of these kinds of measurements, increasing number of research on time series data protection and privacy issue during the mining process is expected.

7. Conclusion

In this paper, we have reviewed research in time series data mining. Different research is focused on one or more problems in time series data mining. However, according to the unique behavior of the time series data, existing research is still inadequate and it is considered as one of the 10 challenging problems in data mining (Yang and Wu, 2006). There is still room for us to further investigate and develop. For example, while most of the research communities have concentrated on the mining tasks, the fundamental problem on how to represent a time series has not yet been fully addressed so far. To represent a time series is essential, because time series data is hard to manipulate in its original structure. The high dimensionality of time series data creates difficulties in applying existing data mining techniques to it. Therefore, defining a more effective and efficient time series representation scheme is of fundamental importance. The framework should also support time series pattern matching, including both whole sequence and subsequence matching, between time series of different lengths in an effective manner. The framework should be compatible to varieties of time series data mining tasks like pattern discovery. In addition, handling multi-attribute time series data, mining on time series data stream and privacy issue are three promising research directions, due to the existence of the system with high computational power.

References

ABfalg, J., Kriegel, H.P., Kroger, P., Kunath, P., Pryakhin, A., Renz, M., 2006. Similarity search on time series based on threshold queries. In: Proceedings of the 10th International Conferece on Extending Database Technology, pp. 276–294.

ABfalg, J., Kriegel, H.P., Kroger, P., Kunath, P., Pryakhin, A., Renz, M., 2008. T-Time: threshold-based data mining on time series. In: Proceedings of the 24th IEEE International Conference on Data Engineering, pp. 1620–1623.

Abonyi, J., Feil, B., Nemeth, S., Arva, P., 2003. Principal component analysis based time series segmentation—application to hierarchical clustering for multivariate process data. In: Proceedings of the IEEE International Conference on Computational Cybernetics, pp. 29–31.

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.

Agrawal, R., Srikant, R., 2000. Privacy-preserving Data Mining. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 439–450.

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

Agrawal, R., Imielinski, T., Swami, A., 1993b. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 207–216.

Agrawal, R., Lin, K.I., Sawhney, H.S., Shim, K., 1995a. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In: Proceedings of the 21st International Conference on Very Large Databases, pp. 490–501.

Agrawal, R., Psaila, G., Wimmers, E.L., Zait, M., 1995b. Querying shapes of histories. In: Proceedings of the 21st International Conference on Very Large Databases, pp. 502–514.

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference Very Large Databases, pp. 487–499.

Ahmad, S., Oliveira, P.C.F., Ahmad, K., 2004a. Summarization of multimodal information. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, vol. 3, pp. 1049–1052.

Ahmad, S., Oliveira, P.C.F., Ahmad, K., 2004a. Summarization of multimodal information. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, vol. 3, pp. 1049–1052.

Ahmad, S., Taskaya-Temizel, T., Ahmad, K., 2004b. Summarizing time series: learning patterns in volatile series. In: Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning, pp. 523–532.

Aref, W.G., Elfeky, M.G., Elmagarmid, A.K., 2004. Incremental, online, and merge mining of partial periodic patterns in time-series databases. IEEE Transactions on Knowledge and Data Engineering *16* (*3*), 332–342.

Arentz, W.A., Hetland, M.L., Olstad, B., 2005. Retrieving musical information based on rhythm and pitch correlations. Journal of New Music Research *34* (*2*), 151–159.

Aris, A., Shneiderman, B., Plaisant, C., Shmueli, G., Jank, W., 2005. Representing unevenly-spaced time series data for visualization and interactive exploration. In: Proceedings of the International Conference on Human–Computer Interaction, pp. 835–846.

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

Azzouzi, M., Nabney, I.T., 1998. Analysing time series structure with Hidden Markov Models. In: Proceedings of the IEEE Conference on Neural Networks and Signal Processing, pp. 402–408.

Bagnall, A., Janacek, G., 2005. Clustering time series with clipped data. Machine Learning 58 (2–3), 151–178.

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

Bao, D.A., 2008. Generalized model for financial time series representation and prediction. Applied Intelligence 29 (1), 1–11.

Bao, D., Yang, Z., 2008. Intelligent stock trading system by turning point confirming and probabilistic reasoning. International Journal of Expert Systems with Applications 34 (1), 620–627.

Bargiela, A., Pedrycz, W., 2003. Recursive information granulation: aggregation and interpretation issues. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics *33* (*1*), 96–112.

Batyrshin, I.Z., Sheremetov, L.B., 2008. Perception-based approach to time series data mining. Applied Soft Computing 8 (3), 1211–1221.
Beckmann, N., Kriegel, H.P., Schneider, R., Seeger, B., 1990. The R*-tree: an efficient and robust access method for points and rectangles. In: Proceedings of the 1990 ACM SIGMOD International Conference on Management of Data, pp. 322–331.
Berberidis, C., Vlahavas, I.P., Aref, W.G., Atallah, M.J., Elmagarmid, A.K., 2002a. On the discovery of weak periodicities in large time series. In: Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 51–61.
Berberidis, C., Walid, A.G., Atallah, M., Vlahavas, I., Elmagarmid, A.K., 2002b. Multiple and partial periodicity mining in time series databases. In: Proceedings of the 15th European Conference on Artificial Intelligence.
Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: AAAI Working Notes of the Knowledge Discovery in Databases Workshop, pp. 359–370.
Berndt, D.J., Clifford, J., 1996. Finding patterns in time series: a dynamic programming approach. Advances in Knowledge Discovery and Data Mining, 229–248.
Bettini, C., Wang, S., Jajodia, S., Lin, J.L., 1998. Discovering frequent event patterns with multiple granularities in time sequences. IEEE Transactions on Knowledge and Data Engineering *10* (*2*), 222–237.
Boyd, S., 1998. TREND: a system for generating intelligent descriptions of time-series data. In: Proceedings of the IEEE International Conference on Intelligent Processing Systems.
Bozkaya, T., Yazdani, N., Ozsoyoglu, Z.M., 1997. Matching and indexing sequences of different lengths. In: Proceedings of the Sixth ACM International Conference on Information and Knowledge Management, pp. 128–135.
Buono, P., Aris, A., Plaisant, C., Khella, A., Shneiderman, B., 2005. Interactive pattern search in time series. In: Proceedings of the Conference on Visualization and Data Analysis, pp. 175–186.
Capitani, P., Ciaccia, P., 2007. Warping the time on data streams. Data and Knowledge Engineering 62 (3), 438–458.
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.
Chan, K.P., Fu, A.C., 1999. Efficient time series matching by wavelets. In: Proceedings of the 15th IEEE International Conference on Data Engineering, pp. 126–133.
Chan, K.P., Fu, A., Yu, C., 2003. Haar wavelets for efficient similarity search of time-series: with and without time warping. IEEE Transactions on Knowledge and Data Engineering *15* (*3*), 685–705.
Chan, P., Mahoney, M., 2005. Modeling multiple time series for anomaly detection. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 90–97.
Chen, J., 2005. Making subsequence time series clustering meaningful. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 114–121.
Chen, J., 2007. Useful clustering outcomes from meaningful time series clustering. In: Proceedings of the Sixth Australasian Conference on Data Mining and Analytics, pp. 101–109.
Chen, L., Ozsu, M.T., Oria V., 2005a. Robust and fast similarity search for moving object trajectories. In: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, pp. 491–502.
Chen, L., Ozsu, M.T., Oria, V., 2005b. Using multi-scale histograms to answer pattern existence and shape match queries. In: Proceedings of the 17th International Conference on Scientific and Statistical Database Management.
Chen, Q. Chen, L., Lian, X., Liu, Y., Yu, J.X., 2007a. Indexable PLA for efficient similarity search. In: Proceedings of the 33rd International Conference on Very Large Databases, pp. 435–446.
Chen, Y., Dong, G., Han, J., Wah, B.W., Wang, J., 2002. Multi-dimensional regression analysis of time-series data streams. In: Proceedings of the 28th International Conference on Very Large Databases, pp. 323–334.
Chen, Y., Nascimento, M.A., Ooi, B.C., Tung, A.K.H., 2007b. SpADe: on shape-based pattern detection in streaming time series. In: Proceedings of IEEE 23rd International Conference on Data Engineering, pp. 786–795.
Chiu, B., Keogh, E., Lonardi, S., 2003. Probabilistic discovery of time series motifs. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 493–498.
Chu, C.S.J., 1995. Time series segmentation: a sliding window approach. Information Sciences 85 (1–3), 147–173.
Chu, K.K.W., Wong, M.H., 1999. Fast time-series searching with scaling and shifting. In: Proceedings of the 18th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 237–248.
Chu, K.W., Lam, S.K., Wong, M.H., 1998. An efficient Hash-based algorithm for sequence data searching. The Computer Journal *41* (*6*), 402–415.
Chu, S., Keogh, E., Hart, D., Pazzani, M., 2002. Iterative deepening dynamic time warping for time series. In: Proceedings of the Second SIAM International Conference on Data Mining.
Chung, F.L., Fu, T.C., Luk, R., Ng, V., 2001. Flexible time series 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.
Chung, F.L., Fu, T.C., Ng, V., Luk, R., 8, 2004. An evolutionary approach to pattern-based time series segmentation. IEEE Transactions on Evolutionary Computation, 471–489.
Cole, R., Shasha, D., Zhao, X., 2005. Fast window correlations over uncooperative time series. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 743–749.
Cotofrei, P., Stoffel, K., 2002. Classification rules+time=temporal rules. In: Proceedings of the 2002 International Conference on Computational Science, pp. 572–581.
Das, G., Gunopulos, D., Mannila, H., 1997. Finding similar time series. In: Proceedings of the First European Symposium on Principles and Practice of Knowledge Discovery in Databases, pp. 88–100.
Das, G., Lin, K.I., Mannila, H., Renganathan, G., Smyth, P., 1998. Rule discovery from time series. In: Proceedings of the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 16–22.
Dasha, P.K., Nayaka, M., Senapatia, M.R., Lee, I.W.C., 2007. Mining for similarities in time series data using wavelet-based feature vectors and neural networks. Engineering Applications of Artificial Intelligence 20 (2), 185–201.
Deligiannakis, A., Kotidis, Y. and Roussopoulos, N., 2004. Compressing historical information in sensor networks. In: Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, pp. 527–538.
Denton, A., 2005. Kernel-density-based clustering of time series subsequences using a continuous random-walk noise model. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 122–129.
Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E., 2008. Querying and mining of time series data: experimental comparison of representations and distance measures. In: Proceedings of the VLDB Endowment, vol. 1(2), pp. 1542–1552.
Dong, X.L., Gu, C.K., Wang, Z.O., 2009. Research on shape-based time series similarity measure. In: Proceedings of the 2006 International Conference on Machine Learning and Cybernetics, pp. 1253–1258.
Dorr, A.H., Denton, A.M., 2009. Establishing relationships among patterns in stock market data. Data and Knowledge Engineering 68 (3), 318–337.
Douglas, D., Peucker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer *10* (*2*), 112–122.
Duan, J., Wang, W., Liu, B., Xue, Y., Zhou, H., Shi, B., 2005. Incorporating with recursive model training in time series clustering. In: Proceedings of the Fifth International Conference on Computer and Information Technology, pp. 105–109.
Duncan, S.R., Bryant, G.F. A, 1996. New algorithm for segmenting data from time series. In: Proceedings of the 35th Conference on Decision and Control, pp. 3123–3128.
Elfeky, M.G., Aref, W.G., Elmagarmid, A.K., 2004. Using convolution to mine obscure periodic patterns in one pass. In: Proceedings of the Ninth International Conference on Extending Database Technology, pp. 605–620.
Elfeky, M.G., Aref, W.G., Elmagarmid, A.K., 2005. Periodicity detection in time series databases. IEEE Transactions on Knowledge and Data Engineering *17* (*7*), 875–887.
Euachongprasit, W., Ratanamahatana, C.A., 2008. Accurate and efficient retrieval of multimedia time series data under uniform scaling and time warping. In: Proceedings of the 12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp. 100–111.
Euliano, N.R., Principe, J.C., 1996. Spatio-temporal self-organizing feature maps. In: Proceedings of the IEEE International Conference on Neural Networks, vol. 4, pp. 1900–1905.
Faloutsos, C., Jagadish, H., Mendelzon, A., Milo, T., 1997. A Signature technique for similarity-based queries. In: Proceedings of the International Conference on Compression and Complexity of Sequences, pp. 2–20.
Faloutsos, C., Ranganathan, M., ManolopoulosY., 1994. Fast subsequence matching in time-series databases. In: Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, pp. 419–429.
Fancourt, C.L., Principe, J.C., 1996. A neighborhood map of competing one step predictors for piecewise segmentation and identification of time series. In: Proceedings of the International Conference on Neural Network, vol. 4, pp. 1906–1911.
Fancourt, C.L., 1997. and Principe, J.C. Competitive principal component analysis for locally stationary time series. IEEE Transactions on Signal Processing *46* (*11*), 3068–3082.
Feng, L., Ju, K., Chon, K.H.A., 2005. Method for segmentation of switching dynamic modes in time series. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics *35* (*5*), 1058–1064.
Fink, E., Pratt, K.B., Gandhi, H.S., 2003. Indexing of time series by major minima and maxima. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, pp. 2332–2335.
Fitzgibbon, L.J., Dowe, D.L., Allison, L., 2002. Change-point estimation using new minimum message length approximations. In: Proceedings of the Seventh Pacific Rim International Conference on Artificial Intelligence: Trends in Artificial Intelligence, pp. 244–254.
Fu, A., Keogh, E., Lau, L., Ratanamahatana, C.A., Wong, C.W., 2008a. Scaling and time warping in time series querying. The VLDB Journal 17 (4), 899–921.
Fu, T.C., Chung, F.L., Kwok, K.Y., Ng, C.M., 2008b. Stock time series visualization based on data point importance. Engineering Applications of Artificial Intelligence 21 (8), 1217–1232.
Fu, T.C., Chung, F.L., Luk, R., Ng, C.M., 2005a. Preventing meaningless stock time series pattern discovery by changing perceptually important point detection,

In: Proceeding of the Second International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1171–1174.
Fu, T.C., Chung, F.L., Luk, R., Ng, C.M., 2008c. Representing financial time series based on data point importance. Engineering Applications of Artificial Intelligence 21 (2), 277–300.
Fu, T.C., Chung, F.L., Luk, R., Ng, C.M., 2007. Stock time series pattern matching: template-based vs. rule-based approaches. Engineering Applications of Artificial Intelligence 20 (3), 347–364.
Fu, T.C., Chung, F.L., Luk, R., Ng, V., 2001. Pattern discovery from stock time series using self-organizing maps. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Temporal Data Mining, pp. 27–37.
Fu, T.C., Chung, F.L., Ng, C.M., 2006. Financial time series segmentation based on specialized binary tree representation. In: Proceedings of the 2006 International Conference on Data Mining, pp. 3–9.
Fu, T.C., Chung, F.L., Tang, P.Y., Luk, R., Ng, C.M., 2005b. Incremental stock time series data delivery and visualization. In: Proceedings of the 14th ACM Conference on Information and Knowledge Management, pp. 279–280.
Gao, L., Wang, X.S., 2002. Continually evaluating similarity-based pattern queries on a streaming time series. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 370–381.
Gao, L., Yao, Z., Wang, X.S., 2002. Evaluating continuous nearest neighbor queries for streaming time series via pre-fetching. In: Proceedings of the 11th ACM International Conference on Information and Knowledge Management, pp. 485–492.
Gavrilov, M., Anguelov, D., Indyk, P., Motwani, R., 2000. Mining the stock market: which measure is best? In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 487–496.
Ge, X., Smyth, P., 2000. Deformable Markov Model templates for time-series pattern matching. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 81–90.
Ge, X.P., 1998. Pattern matching in financial time series data. University of California, Irvine, Final Project Report for. ICS 278.
Geurts, P., 2001. Pattern extraction for time series classification. In: Proceedings of the Fifth European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 115–127.
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.
Golab, L., Ozsu, M.T., 2003. Issues in data stream management. ACM SIGMOD Record 32 (2), 5–14.
Goldin, D., Kanellakis, P., 1995. On similarity queries for time-series data: constraint specification and implementation. In: Proceedings of the First International Conference on Principles and Practice of Constraint Programming, pp. 137–153.
Grass, J., Ziberstein, S., 1996. Anytime algorithm development tools. Sigart Artificial Intelligence 7 (2).
Guimaraes, G., Ultsch, A., 1999. A Method for temporal knowledge conversion. In: Proceedings of the Third International Symposium on Intelligent Data Analysis, pp. 369–382.
Guimaraes, G., Peter, J.H., Penzel, T., Ultsch, A.A., 2001. Method for automated temporal knowledge acquisition applied to sleep-related breathing disorders. Artificial Intelligence in Medicine *23* (3), 211–237.
Guo, X., Liang, X., Li, N., 2007. Automatically recognizing stock patterns using RPCL neural networks. In: Proceedings of the 2007 International Conference on Intelligent Systems and Knowledge Engineering, pp. 997–1004.
Guralnik, V., Srivastava, J., 1999. Event detection from time series data. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 33–42.
Gusfield, D., 1997. Algorithms on Strings, Trees and Sequences. Cambridge University Press.
Guttman, A., 1984. R-trees: a dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, pp. 47–57.
Han, J., Dong, G., Yin, Y., 1999. Efficient mining of partial periodic patterns in time series database. In: Proceedings of the 15th IEEE International Conference on Data Engineering, pp. 106–115.
Han, J., Gong, W., Yin, Y., 1998. Mining segment-wise periodic patterns in time-related databases. In: Proceedings of the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 214–218.
Han, W.S., Lee, J., Moon, Y.S., Jiang, H., 2007. Ranked subsequence matching in time-series databases. In: Proceedings of the 33rd International Conference on Very Large Databases, pp. 423–434.
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.
Hebrail, G., Hugueney, B., 2001. Symbolic representation of long time-series. In: Proceedings of the Applied Stochastic Models and Data Analysis Conference.
Hershberger, J., Snoeyink, J., 1992. Speeding up the Douglas–Peucker line-simplification algorithm. In: Proceedings of the Fifth Symposium on Data Handling, pp. 134–143.
Hetland, M.L., Saetrom, P., 2005. Evolutionary rule mining in time series databases. Machine Learning 58 (2–3), 107–125.
Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., Toivonen, H., 2001. Time series segmentation for context recognition in mobile devices. In: Proceedings of the 2001 IEEE International Conference on Data Mining, pp. 203–210.
Hochheiser, H., Shneiderman, B., 2004. Dynamic query tools for time series data sets, timebox widgets for interactive exploration. Information Visualization *3* (*1*), 1–18.
Hoppner, F., Klawonn, F., 2001. Finding informative rules in interval sequences. In: Proceedings of the Fourth International Symposium on Intelligent Data Analysis, pp. 123–132.
Huang, Y.W., Yu, P.S., 1999. Adaptive query processing for time-series data. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 282–286.
Hugueney, B., Meunier, B.B., 2001. Time-series segmentation and symbolic representation, from process-monitoring to data-mining. In: Proceedings of the Seventh International Conference on Computational Intelligence, Theory and Applications, pp. 118–123.
Indyk, P., Koudas, N., Muthukrishnan, S., 2000. Identifying representative trends in massive time series data sets using sketches. In: Proceedings of the 26th International Conference on Very Large Databases, pp. 363–372.
Jagadish, H.V., Mendelzon, A.O., Milo, T., 1995. Similarity-based queries. In: Proceedings of the 14th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 36–45.
Janacek, G.J., Bagnall, A.J., Powell, M.A., 2005. Likelihood ratio distance measure for the similarity between the Fourier transform of time series. In: Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 737–743.
Jia, P., He, H., Sun, T., 2008. Error restricted piecewise linear representation of time series based on special points. In: Proceedings of the Seventh World Congress on Intelligent Control and Automation, pp. 2059–2064.
Jiang, J., Zhang, Z., Wang, H., 2007. A New segmentation algorithm to stock time series based on PIP approach. In: Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, pp. 5609–5612.
Jin, X., Lu, Y., Shi, C., 2002. Distribution discovery: local analysis of temporal rules. In: Proceedings of the Sixth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 469–480.
Jonsson, H.A., Badal, Z., 1997. Using signature files for querying time-series data. In: Proceedings of the First European Symposium on Principles and Practice of Knowledge Discovery in Databases, pp. 211–220.
Kacprzyk, J., Wilbik, A., Zadrozny, S., 2008. Linguistic summarization of time series using a fuzzy quantifier driven aggregation. Fuzzy Sets and Systems 159 (12), 1485–1499.
Kadous, M.W., Sammut, C., 2005. Classification of multivariate time series and structured data using constructive induction. Machine Learning 58 (2–3), 179–216.
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.
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.
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.
Kawagoe, K., Ueda, T., 2002. A Similarity search method of time series data with combination of Fourier and wavelet transforms. In: Proceedings of the Ninth IEEE International Symposium on Temporal Representation and Reasoning, pp. 86–93.
Keogh, E., 1997a. A fast and robust method for pattern matching in time series databases. In: Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence, pp. 578–584.
Keogh, E., 1997b. 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.
Keogh, E., 2002. Exact indexing of dynamic time warping. In: Proceedings of the 28th International Conference on Very Large Databases, pp. 406–417.
Keogh, E., Kasetty, S., 2002. On the need for time series 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.
Keogh, E., Pazzani, M., 1998. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In: Proceedings of the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 239–341.
Keogh, E., Pazzani, M., 1999. Relevance feedback retrieval of time series data. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 183–190.
Keogh, E., Pazzani, M., 2000a. A Simple dimensionality reduction technique for fast similarity search in large time series databases. In: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 122–133.
Keogh, E., Pazzani, M., 2000b. Scaling up dynamic time warping for datamining applications. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 285–289.
Keogh, E., Pazzani, M., 2001. Derivative dynamic time warping. In: Proceedings of the First SIAM International Conference on Data Mining.

Keogh, E., Smyth, P.A., 2001. Probabilistic approach to fast pattern matching in time series databases. In: Proceedings of the Third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1997, pp. 24–30.
Keogh, E., Chakrabarti, K., Mehrotra, S., Pazzani, M., 2001a. Locally adaptive dimensionality reduction for indexing large time series databases. In: Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data, pp. 151–163.
Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S., 2000. Dimensionality reduction for fast similarity search in large time series databases. Journal of Knowledge and Information Systems *3* (*3*), 263–286.
Keogh, E., Chu, S., Pazzani, M., 2001b. Ensemble-index: a new approach to indexing large databases. In: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 117–125.
Keogh, E., Chu, S., Hart, D., Pazzani, M., 2001c. An online algorithm for segmenting time series. In: Proceedings of the 2001 IEEE International Conference on Data Mining, pp. 289–296.
Keogh, E., Hochheiser, H., Shneiderman, B., 2002a. An augmented visual query mechanism for finding patterns in time series data. In: Proceedings of the Fifth International Conference on Flexible Query Answering Systems, pp. 240–250.
Keogh, E., Lin, J., Fu, A., 2005. HOT SAX: efficiently finding the most unusual time series subsequence. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 226–233.
Keogh, E., Lin, J., Truppel, W., 2003. Clustering of time series subsequences is meaningless: implications for previous and future research. In: Proceedings of the Third IEEE International Conference on Data Mining, pp. 115–122.
Keogh, E., Lin, J., Fu, A., Herle, H.V., 2006. Finding unusual medical time-series subsequences: algorithms and applications. IEEE Transactions on Information Technology in Biomedicine *10* (*3*), 429–439.
Keogh, E., Lin, J., Lee, S.H., Herle, H.V., 2007a. Finding the most unusual time series subsequence: algorithms and applications. Knowledge and Information Systems 11 (1), 1–27.
Keogh, E., Lonardi, S., Chiu, Y.C., 2002b. Finding surprising patterns in a time series database in linear time and space. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 550–556.
Keogh, E., Lonardi, S., Ratanamahatana, C.A., 2004. Towards parameter-free data mining. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 206–215.
Keogh, E., Lonardi, S., Ratanamahatana, C.A., Wei, L., Lee, S.H., Handley, J., 2007b. Compression-based data mining of sequential data. Data Mining and Knowledge Discovery 14 (1), 99–129.
Kim, M.S., Kim, S.W., Shin, M., 2005. Optimization of subsequence matching under time warping in time-series databases. In: Proceedings of the 2005 ACM Symposium on Applied Computing, pp. 581–586.
Kim, S.W., Jeong, B.S., 2007. Performance bottleneck of subsequence matching in time-series databases: observation, solution, and performance evaluation. Information Sciences 177 (22), 4841–4858.
Kim, S.W., Park, S.H., Chu, W.W., 2001. An index-based approach for similarity search supporting time warping in large sequence databases. In: Proceedings of the 17th IEEE International Conference on Data Engineering, pp. 607–614.
Kim, S.W., Yoon, J., Park, S., Kim, T.H., 2002. Shape-based retrieval of similar subsequences in time-series databases. In: Proceedings of the 2002 ACM Symposium on Applied Computing, pp. 438–445.
Kim, Y.I., Park, Y., Chun, J., 1996. A Dynamic indexing structure for searching time-series pattern. In: Proceedings of the 20th Computer Software and Applications Conference, pp. 270–275.
Kim, Y.I., Ryu, K.H., Park, T., 1994. Algorithms of improved multidimensional dynamic index for the time-series pattern. In: Proceedings of the First KIPS Spring Conference, vol. 1(1).
Kohonen, T., 1995. Self-Oranizing Maps. Springer, Berlin.
Korn, F., Jagaciish, H.V., Faloutsos, C., 1997. Efficiently supporting ad hoc queries in large data sets of time sequences. In: Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data, pp. 289–300.
Kruskall, J.B., Liberman, M., 1983. The symmetric time warping algorithm: from continuous to discrete, Time Warps, String Edits and Macromolecules. *Addison-Wesley*.
Kumar, N., Lolla, N., Keogh, E., Lonardi, S., Ratanamahatana, C.A., 2005. Time-series bitmaps A practical visualization tool for working with large time series databases. In: Proceedings of the Fifth SIAM International Conference on Data Mining.
Kuo, S.C., Li, S.T., Cheng, Y.C., Ho, M.H., 2004. Knowledge discovery with SOM networks in financial investment strategy. In: Proceedings of the 4th International Conference on Hybrid Intelligent Systems, pp. 98–103.
Lam, S.K., Wong, M.H.A., 1998. Fast projection algorithm for sequence data searching. Data and Knowledge Engineering *28* (*3*), 321–339.
Lang, W., Morse, M., Patel, J., 2010. Dictionary-based compression for long time-series similarity. In: IEEE Transactions on Knowledge and Data Engineering 22 (11), 1609–1622.
Last, M., Klein, Y., Kandel, A., 2001. Knowledge discovery in time series databases. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics *31* (*1*), 160–169.
Latecki, L.J., Megalooikonomou, V., Wang, Q., Lakaemper, R., Ratanamahatana, C.A., Keogh, E., 2005. Partial elastic matching of time series. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 701–704.
Lavrenko, V., Schmill, M., Lawrie, D., Ogilvie, P., Jensen, D., Allan, J., 2000. Mining of concurrent text and time series. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Text Mining, pp. 37–44.
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.
Lee, S., Kwon, D., Lee, S., 2003. Dimensionality reduction for indexing time series based on the minimum distance. Journal of Information Science and Engineering 19, 697–711.
Lei, H., Govindaraju, V., 2004. Regression time warping for similarity measure of sequence. In: Proceedings of the Fourth International Conference on Computer and Information Technology, pp. 826–830.
Leigh, W., Modani, N., Purvis, R., Roberts, T., 2002. Stock market trading rule discovery using technical charting heuristics. Expert Systems with Applications 23, 155–159.
Leng, M., Lai, X., Tan, G., Xu, X., 2009. Time series representation for anomaly detection. In: Proceedings of the Second IEEE International Conference on Computer Science and Information Technology, pp. 628–632.
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.
Lesch, R., Caille, Y., Lowe, D., 1999. Component analysis in financial time series. In: Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering, pp. 183–190.
Li, B., Tan, L., Zhang, J., Zhuang, Z., 2000a. Using fuzzy neural network clustering algorithm in the symbolization of time series. In: Proceedings of the 2000 IEEE Asia Pacific Conference on Circuits and Systems, pp. 379–382.
Li, C., Yu, P.S., Castelli, V., 1998. MALM: a framework for mining sequence database at multiple abstraction levels. In: Proceedings of the Seventh ACM International Conference on Information and Knowledge Management, pp. 267–272.
Li, C.S., Yu, P.S., Castelli, V., 1996. HierarchyScan: a hierarchical similarity search algorithm for databases of long sequences. In: Proceedings of the 12th IEEE International Conference on Data Engineering, pp. 546–553.
Li, Q., Lopez, I.F.V., Moon, B., 2004. Skyline index for time series data. IEEE Transactions on Knowledge and Data Engineering *16* (*6*), 669–684.
Li, Y., Wang, X.S., Jajodia, S., 2000b. Discovering temporal patterns in multiple granularities. In: the International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining, pp. 5–19.
Lian, X., Chen, L., 2008. Efficient similarity search over future stream time series. IEEE Transactions on Knowledge and Data Engineering 20 (1), 40–54.
Lian, X., Chen, L., Yu, J.X., 2008. Pattern matching over cloaked time series. In: Proceedings of the 24th IEEE International Conference on Data Engineering, pp. 1462–1464.
Lian, X., Chen, L., Yu, J.X., Han, J., Ma, J., 2009. Multiscale representations for fast pattern matching in stream time series. IEEE Transactions on Knowledge and Data Engineering 21 (4), 568–581.
Lim, H.S., Whang, K.Y., Moon, Y.S., 2008. Similar sequence matching supporting variable-length and variable-tolerance continuous queries on time-series data stream. Information Sciences 178 (6), 1461–1478.
Lim, S.H., Park, H., Kim, S.W., 2007. Using multiple indexes for efficient subsequence matching in time-series databases. Information Sciences 177 (24), 5691–5706.
Lin, J., Li, Y., 2009. Finding structural similarity in time series data using bag-of-patterns representation. In: Proceedings of the 21st International Conference on Scientific and Statistical Database Management, pp. 461–477.
Lin, J., Keogh, E., Lonardi, S., 2005a. Visualizing and discovering non-trivial patterns in large time series databases. Information Visualization *4* (*2*), 61–82.
Lin, J., Keogh, E., Lonardi, S., Chiu, B., 2003. A Symbolic representation of time series, 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.
Lin, J., Keogh, E., Lonardi, S., Patel, P., 2002. Finding motifs in time series. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2nd Workshop on Temporal Data Mining, pp. 53–68.
Lin, J., Keogh, E., Wei, L., Lonardi, S., 2007. Experiencing SAX: a novel symbolic representation of time series. Data Mining and Knowledge Discovery 15 (2), 107–144.
Lin, J., Vlachos, M., Keogh, E., Gunopulos, D., 2004. Iterative incremental clustering of time series. In: Proceedings of the Ninth International Conference on Extending Database Technology, pp. 106–122.
Lin, J., Vlachos, M., Keogh, E., Gunopulos, D., Liu, J.W., Yu, S.J., Le, J.J., 2005b. A MPAA-based iterative clustering algorithm augmented by nearest neighbors search for time-series data streams. In: Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 333–342.
Lin, L., Risch, T., 1998. Querying continuous time sequences. In: Proceedings of the 24th International Conference on Very Large Databases, pp. 170–181.
Liu, J.N.K., Kwong, R.W.M., Bo, F., 2004. Chart patterns recognition and forecast using wavelet and radial basis function network. In: Proceedings of the Eighth International Conference on Knowledge-Based Intelligent Information and Engineering Systems, pp. 564–571.
Liu, X., Lin, Z., Wang, H., 2008. Novel online methods for time series segmentation. IEEE Transactions on Knowledge and Data Engineering 20 (12), 1616–1626.
Loh, W.K., Kim, S.W., 2001. A Subsequence matching algorithm supporting moving average transform of arbitrary order in time-series databases using

index interpolation. In: Proceedings of the Australian Database Conference, pp. 37–44.
Loh, W.K., Moon, Y.S., Srivastava, J., 2010. Distortion-free predictive streaming time-series matching. Information Sciences 180 (8), 1458–1476.
Lonardi, S., Lin, J., Keogh, E., Chiu, Y.C., 2006. Efficient discovery of unusual patterns in time series. New Generation Computing 25 (1), 61–93.
Lu, H., Han, J., Feng, L., 1998. Stock movement prediction and N-dimensional inter-transaction association rules. In: Proceedings of the Third ACM SIGMOD International Conference on Management of Data Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 1–7.
Ma, J., Perkins, S., 2003. Online novelty detection on temporal sequences. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 613–618.
Man, P., Wong, M.H., 2001. Efficient and robust feature extraction and pattern matching of time series by a lattice structure. In: Proceedings of the 10th ACM International Conference on Information and Knowledge Management, pp. 271–278.
Megalooikonomou, V., Li, G., Wang, Q., 2004. A Dimensionality reduction technique for efficient similarity analysis of time series databases. In: Proceedings of the 13th ACM International Conference on Information and Knowledge Management, pp. 160–161.
Megalooikonomou, V., Wang, Q., Li, G., Faloutsos, C., 2005. A Multiresolution symbolic representation of time series. In: Proceedings of the 21st IEEE International Conference on Data Engineering, pp. 668–679.
Miller, F.W., Keleher, P. and Tripathi, S.K., 1998. General data streaming. In: Proceedings of the 19th IEEE Real-Time Systems Symposiums, pp. 232–250.
Minnen, D., Isbell, C.L., Essa, I.A., Starner, T., 2007a. Detecting subdimensional motifs: an efficient algorithm for generalized multivariate pattern discovery. In: Proceedings of the Seventh IEEE International Conference on Data Mining, pp. 601–606.
Minnen, D., Isbell, C.L., Essa, I.A., Starner, T., 2007b. Discovering multivariate motifs using subsequence density estimation and greedy mixture learning. In: Proceedings of the 22nd AAAI Conference on Artificial Intelligence, pp. 615–620.
Moller-Levet, C.S., Klawonn, F., Cho, K.H., Wolkenhauer, O., 2003. Fuzzy clustering of short time-series and unevenly distributed sampling points. In: Proceedings of the Fifth International Symposium on Intelligent Data Analysis, pp. 330–340.
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.
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.
Morchen, F., 2003. Time series feature extraction for data mining using DWT and DFT. University of Marburg, Department of Mathematics and Computer Science, Technical Report no. 33.
Morchen, F. and Ultsch, A., 2005. Optimizing time series discretization for knowledge discovery. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 660–665.
Morchen, F., Ultsch, A., Hoos, O., 2005. Extracting interpretable muscle activation patterns with time series knowledge mining. International Journal of Knowledge-Based+Intelligent Engineering Systems.
Morinaka, Y., Yoshikawa, M., Amagasa, T., Uemura, S., 2001. The L-index: an indexing structure for efficient subsequence matching in time sequence databases. In: Proceedings of the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 51–60.
Morrill, J.P., 1998. Distributed recognition of patterns in time series data. Communications of the ACM *41* (*5*), 45–51.
Morse, M.D., Patel, J.M., 2007. An efficient and accurate method for evaluating time series similarity. In: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, pp. 569–580.
Motoyoshi, M., Miura, T., Watanabe, K., 2002. Mining temporal classes from time series data. In: Proceedings of the 11th ACM International Conference on Information and Knowledge Management, pp. 493–498.
Mueen, A., Keogh, E., Bigdely-ShamloN., 2009. Finding time series motifs in disk-resident data. In: Proceedings of the 2009 IEEE International Conference on Data Mining, pp. 367–376.
Nin, J., Torra, V., 2009. Towards the evaluation of time series protection methods. Information Sciences 179 (11), 1663–1677.
Oates, T., 1999. Identifying distinctive subsequences in multivariate time series by clustering. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 322–326.
Oates, T., Firoiu, L., Cohen, P.R., 1999. Clustering time series with Hidden Markov Models and dynamic time warping. In: Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Sequence Learning.
Ogras, Y., Ferhatosmanoglu, H., 2006. Online summarization of dynamic time series data. The International Journal on Very Large Data Bases *15* (*1*), 84–98.
Ohsaki, M., Sato, Y., Yokoi, H., Yamaguchi, T., 2003. A Rule discovery support system for sequential medical data in the case study of a chronic hepatitis data set. In: the 14th European Conference on Machine Learning/the Seventh European Conference on Principles and Practice in Knowledge Discovery in Databases Discovery Challenge Workshop, pp. 154–165.
Oliver, J.J., Forbes, C.S., 1997. Bayesian approaches to segmenting a simple time series. In: Proceedings of the Econometric Society Australasian Meeting.
Oliver, J.J., Bexter, R.A., Wallace, C.S., 1998. Minimum message length segmentation. In: Proceedings of the Second Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 222–233.
Palpanas, T., Vlachos, M., Keogh, E., Gunopulos, D., 2008. Streaming time series summarization using user-defined amnesic functions. IEEE Transactions on Knowledge and Data Engineering 20 (7), 992–1006.
Plant, C., Wohlschlager, A.M., Zherdin, A., 2009. Interaction-based clustering of multivariate time series. In: Proceedings of the 2009 IEEE International Conference on Data Mining, pp. 914–919.
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.
Papadimitriou, S., Sun, J., Faloutsos, C., 2005. Streaming pattern discovery in multiple time-series. In: Proceedings of the 31st International Conference on Very Large Databases, pp. 697–708.
Papadimitriou, S., Yu, P., 2006. Optimal multi-scale patterns in time series streams. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pp. 647–658.
Park, S., Chu, W., Yoon, J., Hsu, C., 2000. Efficient searches for similar subsequences of different lengths in sequence databases. In: Proceedings of the 16th IEEE International Conference on Data Engineering, pp. 23–32.
Park, S., Kim, S., Chu, W.W., 2001a. Segment-based approach for subsequence searches in sequence databases. In: Proceedings of the 16th ACM Symposium on Applied Computing, pp. 248–252.
Park, S., Kim, S.W., Cho, J.S., Padmanabhan, S., 2001b. Prefix-querying: an approach for effective subsequence matching under time warping in sequence databases. In: Proceedings of the 10th ACM International Conference on Information and Knowledge Management, pp. 255–262.
Park, S., Lee, D., Chu, W., 1999. Fast retrieval of similar subsequences in long sequence databases. In: Proceedings of the Third IEEE Knowledge and Data Engineering Exchange Workshop, pp. 60–67.
Patel, P., Keogh, E., Lin, J., Lonardi, S., 2002. Mining motifs in massive time series databases. In: Proceedings of the 2002 IEEE International Conference on Data Mining, pp. 370–377.
Perng, C.S., Wang, H., Zhang, R., Parker, D., 2000. Landmarks: a new model for similarity-based pattern querying in time series databases. In: Proceedings of the 16th IEEE International Conference on Data Engineering, pp. 33–42.
Policker, S., Geva, A.B., 2000. Nonstationary time series analysis by temporal clustering. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics *30* (*2*), 339–343.
Popivanov, I., Miller, J., 2002. Similarity search over time-series data using wavelets. In: Proceedings of the 18th IEEE International Conference on Data Engineering, pp. 212–224.
Povinelli, J., Feng, X., 1999. Data mining of multiple nonstationary time series. In: Proceedings of Artificial Neural Networks in Engineering, pp. 511–516.
Povinelli, R.J., Johnson, M.T., Lindgren, A.C., Ye, J., 2004. Time series classification using Gaussian mixture models of reconstructed phase spaces. IEEE Transactions on Knowledge and Data Engineering *16* (*6*), 779–783.
Pratt, B., Fink, E., 2002. Search for patterns in compressed time series. International Journal of Image and Graphics *2* (*1*), 89–106.
Qu, Y., Wang, C., Wang, S., 1998. Supporting fast search in time series for movement patterns in multiple scales. In: Proceedings of the Seventh ACM International Conference on Information and Knowledge Management, pp. 251–258.
Rafiei, D., 1999. On similarity-based queries for time series data. In: Proceedings of the 15th IEEE International Conference on Data Engineering, pp. 410–417.
Rafiei, D., Mendelzon, A., 2000. Querying time series data based on similarity. IEEE Transactions on Knowledge and Data Engineering *12* (*5*), 675–693.
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.
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.
Ratanamahatana, C.A., Keogh, E., Bagnall, A.J., Lonardi, S. A, 2005. Novel bit level 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.
Reiter, E., Sripada, S., Hunter, J., Yu, J., Davy, I., 2005. Choosing words in computer-generated weather forecasts. Artificial Intelligence 167 (1–2), 137–169.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.
Rodrigues, P.P., Gama, J., Pedroso, J.P., 2008. Hierarchical clustering of time series data streams. IEEE Transactions on Knowledge and Data Engineering 20 (5), 615–627.
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.
Ruengronghirunya, P., Niennattrakul, V., Ratanamahatana, C.A., 2009. Speeding up similarity search on a large time series data set under time warping distance. In: Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp. 981–988.
Ruspini, E.H., Zwir, I.S., 1999. Automated qualitative description of measurements. In: Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference.

Sakurai, Y., Yoshikawa, M., Faloutsos, C., 2005. FTW: fast similarity search under the time warping distance. In: Proceedings of the 24th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, pp. 326–337.

Salvador, S., Chan, P., 2004. FastDTW: toward accurate dynamic time warping in linear time and space. In: Prcoeedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Mining Temporal and Sequential Data, pp. 70–80.

Sarker, B.K., Mori, T., Uehara, K., 2003. Parallel algorithms for mining association rules in time series data. In: Proceedings of the International Symposium on Parallel and Distributed Processing and Applications, pp. 273–284.

Schreck, Tekusova, Kohlhammer, T., Fellner D., J., 2007. Trajectory-based visual analysis of large financial time series data. ACM SIGKDD Explorations Newsletter, Special Issue on Visual Analytics 9 (2), 30–37.

Sellis T., Roussopoulos, N., Faloutsos, C., 1987. The R+tree: a dynamic index for multidimensional objects. In: Proceedings of the 13th International Conference on Varge Large Data Bases, pp. 507–518.

Shahabi, C., Tian, X., Zhao, W., 2000. TSA-tree: a wavelet-based approach to improve the efficiency of multi-level surprise and trend queries on time-series data. In: Proceedings of the 12th International Conference on Scientific and Statistical Database Management, pp. 55.

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.

Shieh, J., Keogh, E., 2009. iSAX: disk-aware mining and indexing of massive time series data sets. Data Mining and Knowledge Discovery 19 (1), 24–57.

Shou, Y., Mamoulis, N., Cheung, D.W., 2005. Fast and exact warping of time series using adaptive segmental approximations. Machine Learning 58 (2–3), 231–267.

Simon, G., Lee, J.A., Verleysen, M., 2006. Unfolding preprocessing for meaningful time series clustering. Neural Networks *19* (*6*), 877–888.

Singh, S., Stuart, E., 1998. A Pattern matching tool for time-series forecasting. In: Proceedings of the 14th International Conference on Pattern Recognition, pp. 103–105.

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

Sripada, S., Reiter, E., Hunter, J., Yu, J., Davy, I., 2001. Modelling the task of summarising time series data using KA techniques. In: Proceedings of the International Conference on Knowledge Based Systems and Applied Artifical Intelligence, pp. 183–196.

Sripada, S.G., Reiter, E., Hunter, J., Yu, J., 2003. Generating english summaries of time series data using the Gricean Maxims. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 187–196.

Srivastava, A.N., Weigend, A.S., 1996. Improved time series segmentation using gated experts with simulated annealing. In: Proceedings of the IEEE International Conference on Neural Networks, pp. 1883–1888.

Srivastava, A.N., Su, R., Weigend, A.S., 1999. Data mining for features using scale-sensitive gated experts. IEEE Transactions on Pattern Analysis and Machine Intelligence *21* (*12*), 1268–1279.

Steinbach, M., Tan, P.N., Kumar, V., Klooster, S., Potter, C., 2003. Discovery of climate indices using clustering. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 446–455.

Struzik, R., Siebes, A., 1999. The Haar Wavelet Transform in the time series similarity paradigm. In: Proceedings of the Third European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 12–22.

Struzik, Z.R., Siebes, A.P.J.M., 1998. Wavelet transform in similarity paradigm. In: Proceedings of the Second Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 295–309.

Takashi, S., Tatsuya, H., Yasuo, K., 2009. Causality quantification and its applications: structuring and modeling of multivariate time series. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 787–796.

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.

Tang, H., Liao, S.S., 2008. Discovering original motifs with different lengths from time series. Knowledge-Based Systems 21 (7), 666–671.

Taskaya, T., Ahmad, K., 2003. Bimodal visualisation: a financial trading case study. In: Proceedings of the Seventh International Conference on Information Visualisation, pp. 230–236.

Tatavarty, G., Bhatnagar, R., Young, B., 2007. Discovery of temporal dependencies between frequent patterns in multivariate time series. In: Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, pp. 688–696.

Ting, J., Fu, T.C., Chung, F.L., 2006. Mining of stock data: intra- and inter-stock pattern associative classification. In: Proceedings of the 2006 International Conference on Data Mining, pp. 30–36.

Udechukwu, A., Barker, K., Alhajj, R., 2004. Discovering all frequent trends in time series. In: Proceedings of the 2004 Winter International Symposium on Information and Communication Technologies, pp. 1–6.

Ultsch, A., 1999. Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series. Kohonen Maps, 33–46.

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

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

Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., Keogh, E., 2003. Indexing multi-dimensional time-series with support for multiple distance measures. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 216–225.

Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., Keogh, E., 2006. Indexing multidimensional time-series. The International Journal on Very Large Databases *15* (*1*), 1–20.

Vlachos, M., Meek, C., Vagena, Z., Gunopulos, D., 2004. Identifying similarities, periodicities and bursts for online search queries. In: Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, pp. 131–142.

Vlachos, M., Vagena, Z., Castelli, V., Yu, P.S., 2005a. A Multi-metric index for Euclidean and periodic matching. In: Proceedings of the Ninth European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 355–367.

Vlachos, M., Wu, K.L., Chen, S.K., Yu, P.S., 2005b. Fast burst correlation of financial data. In: Proceedings of the Ninth European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 368–379.

Vlachos, M., Yu, P., Castelli, V., 2005c. On periodicity detection and structural periodic similarity. In: Proceedings of the Fifth SIAM International Conference on Data Mining.

Wang, C., Wang, S., 2000. Supporting subseries nearest neighbor search via approximation. In: Proceedings of the Ninth ACM International Conference on Information and Knowledge Management, pp. 314–321.

Wang, H., Perng, C.S., 2001. The S2-tree: an index structure for subsequence matching of spatial objects. In: Proceedings of the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 312–323.

Wang, H., Wang, W., Yang, J., Yu, P.S., 2002. Clustering by pattern similarity in large data sets. In: Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data, pp. 394–405.

Wang, S., Chung, F.L., Deng, Z., 2005a. Why did time series subsequence clustering disguise. International Journal of Information Acquisition 2 (3), 259–266.

Wang, X., Smith, K.A., Hyndman, R.J., 2005b. Dimension reduction for clustering time series using global characteristics. In: Proceedings of the International Conference on Computational Science, pp. 792–795.

Wang, X, Wirth, A., Wang, L., 2007. Structure-based statistical features and multivariate time series clustering. In: Proceedings of the 2007 IEEE International Conference on Data Mining, pp. 351–360.

Wang, Z.J., Willett, P., 2004. Joint segmentation and classification of time series using class-specific features. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics *34* (*2*), 1056–1067.

Weber, M., Alexa, M., Muller, W., 2001. Visualizing time series on spirals. In: Proceedings of the IEEE Symposium on Information Visualization, pp. 7–14.

Wei, L., Keogh, E., 2006. Semi-supervised time series classification. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 748–753.

Wei, L., Keogh, E., Xi, X.,2006. SAXually explicit images: finding unusual shapes. In: Proceedings of the 2006 IEEE International Conference on Data Mining, pp. 711–720.

Wei, L., Keogh, E., Herle, H.V., Mafra-Neto, A., Abbott, R.J., 2007. Efficient query filtering for streaming time series with applications to semisupervised learning of time series classifiers. Knowledge and Information Systems 11 (3), 313–344.

Wei, L., Kumar, N., Lolla, V., Keogh, E., Lonardi, S., Ratanamahatana, C.A., 2005. Assumption-free anomaly detection in time series. In: Proceedings of the 17th International Conference on Statistical and Scientific Database Management, pp. 237–240.

Wu, H., Salzberg, B., Zhang, D., 2004. Online event-driven subsequence matching over financial data streams. In: Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, pp. 23–34.

Wu, H., Salzberg, B., Sharp, G., Jiang, S., Shirato, H., Kaeli, D., 2005. Subsequence matching on structured time series data. In: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, pp. 682–693.

Wu, L., Faloutsos, C., Sycara, K., Payne, T.R., 2000a. FALCON: feedback adaptive loop for content-based retrieval. In: Proceedings of the 26th International Conference on Very Large Databases, pp. 297–306.

Wu, Y., Agrawal, D., El Abbadi, A., 2000b. A Comparison of DFT and DWT based similarity search in time-series databases. In: Proceedings of the Ninth ACM International Conference on Information and Knowledge Management, pp. 488–495.

Xi, X., Keogh, E., Shelton, C., Wei, L., Ratanamahatana, C.A., 2006. Fast time series classification using numerosity reduction. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 1033–1040.

Xiong, Y., Yeung, D.Y., 2004. Time series clustering with ARMA mixtures. Pattern Recognition *37* (*8*), 1675–1689.

Yamanishi, K., Takeuchi, J., 2002. A Unifying framework for detecting outliers and change points from non-stationary time series data. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 676–681.

Yang, K., Shahabi, C., 2005a. A Multilevel distance-based index structure for multivariate time series. In: Proceedings of the 12th IEEE International Symposium on Temporal Representation and Reasoning, pp. 65–73.

Yang, K., Shahabi, C., 2005b. On the stationarity of multivariate time series for correlation-based data analysis. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 805–808.

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

Yang, O., Jia, W., Zhou, P., Meng, X., 1999. A New approach to transforming time series into symbolic sequences. In: Proceedings of the First Joint Conference between the Biomedical Engineering Society and Engineers in Medicine and Biology, pp. 974.

Yang, Q., Wu, X., 2006. 10 Challenging problems in data mining research. International Journal of Information Technology and Decision Making 5 (4), 597–604.

Yang, Z., Zhao, G., 1998. Application of symbolic techniques in detecting determinism in time series. In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 20(5), pp. 2670–2673.

Yankov, D., Keogh, E., Rebbapragada, U., 2008. Disk aware discord discovery: finding unusual time series in terabyte sized data sets. Knowledge and Information Systems 7 (1), 241–262.

Yankov, D., Keogh, E., Lonardi, S., Fu, A., 2005. Dot plots for time series analysis. In: Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence, pp. 159–168.

Yankov, D., Keogh, E., Medina, J., Chiu, B., Zordan, V., 2007. Detecting time series motifs under uniform scaling. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 844–853.

Yi, B., Faloutsos, C., 2000. Fast time sequence indexing for arbitrary Lp norms. In: Proceedings of the 26th International Conference on Very Large Data Bases, pp. 385–394.

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.

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.

Yin, J., Yang, Q., 2005. Integrating Hidden Markov Models and spectral analysis for sensory time series clustering. In: Proceedings of the Fifth IEEE International Conference on Data Mining, pp. 506–513.

Yoon, H., Yang, K., Shahabi, C., 2005. Feature subset selection and feature ranking for multivariate time series. IEEE Transactions on Knowledge and Data Engineering, Special Issue on Intelligent Data Preparation *17* (*9*), 1186–1198.

Yu, J., Reiter, E. Hunter, J., Sripada, S.G., 2004. A New architecture for summarising time series data. In: Proceedings of the Third International Conference on Natural Language Generation, pp. 47–50.

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.

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

Zhao, Y., Zhang, S., 2006. Generalized dimension-reduction framework for recent-biased time series analysis. IEEE Transactions on Knowledge and Data Engineering *18* (*2*), 231–244.

Zhao, Y., Zhang, C., Zhang, S., 2006. Enhancing DWT for recent-biased dimension reduction of time series data. In: Proceedings of the Australian Conference on Artificial Intelligence, pp. 1048–1053.

Zhou, M., Wong, M.H.A., 2005. Segment-wise time warping method for time scaling searching. Information Sciences 173 (1–3), 227–254.

Zhu, Y., Shasha, D., 2003. Warping indexes with envelope transforms for query by humming. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 181–192.

Zhu, Y., Fu, Y., Fu, H., 2008. On privacy in time series data mining. In: Proceedings of the 12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp. 479–493.

Zwir, I., Enrique, E.H., 1999. Qualitative object description: initial reports of the exploration of the frontier. In: Proceedings of Joint EUROFUSE-SIC99 International Conference, pp. 485–490.