ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 7, January 2016

AN IMPROVED TIME SEQUENCE PATTERN MATCHING TECHNIQUE BY USING HYBRIDIZED DCT-SVD APPROACH

R.Rosario Arockia Raj[1], Ms.K.Devipriya[2]
1.PG Student, II ME CSE, Sri Krishna College of Engineering and Technology, Coimbatore
2.Assitant Professor, Department of CSE, Sri Krishna College of Engineering and Technology, Coimbatore

ABSTRACT

Sequence pattern matching in the high dimensional data plays a critical role where the appropriate retrieval of data would be more difficult. To improve the accuracy of the sequence pattern matching approach in the existing work, DCT and DFT based MBR transformation is introduced which would convert the high dimensional MBR into low dimensional MBR directly where the accurate matching would be performed in the less time. The DCT based approach is proved as better approach than the DFT based approach. DCT approach is based on the energy compression technique which might lead to an more computational complexity. This problem can be resolved in the proposed methodology by introducing the hybridized DCT-SVD approach where the SVD would select the most optimal energy efficient component for every block. It will take advantages of DCT first, and use SVD only for the blocks that DCT does not compact energy well. The proposed approach provides better result than the existing approach in terms of improved accuracy and performance evaluation.

Introduction

Time series data

A time series is a sequence of data points, typically consisting of successive measurements made over a time interval. Time series are very frequently plotted via line charts. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, intelligent transport and trajectory forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

Methods for time series analysis

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and recently wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In time domain, correlation analyses can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in frequency domain.

Exploratory analysis

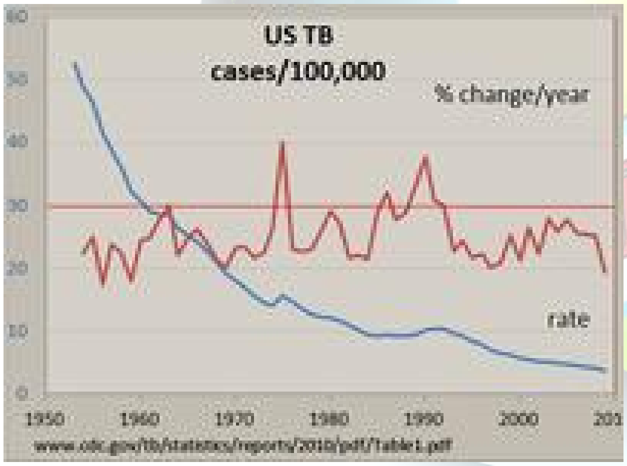

Fig 1.1. Exploratory analysis

The clearest way to examine a regular time series manually is with a line chart such as the one shown for tuberculosis in the United States, made with a spreadsheet program. The number of cases was standardized to a rate per 100,000 and the percent change per year in this rate was calculated. The nearly steadily dropping line shows that the TB incidence was decreasing in most years, but the percent change in this rate varied by as much as +/- 10%, with 'surges' in 1975 and around the early 1990s. The use of both vertical axes allows the comparison of two time series in one graphic. Other techniques include:

- Autocorrelation analysis to examine serial dependence.
- Spectral analysis to examine cyclic behaviour which need not be related to seasonality. For example, sun spot activity varies over 11 year cycles. Other common examples include celestial phenomena, wheather patterns, neural activity, commodity prices, and economic activity.
- Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series.

SEQUENCE MATCHING TECHNIQUE

One approach common to many similar sequence matching techniques is to construct minimum bounding rectangles (MBRs) and use a multidimensional index structure like the R⁄-tree. MBRs are used to reduce the number of data window sequences stored in the index or the number of query window sequences used to search the index. All these techniques use lower-dimensional transformation to reduce high-dimensional sequences to low-dimensional sequences. This transformation is needed to avoid the curse of high dimensionality. Besides, MBRs reduce the required index storage space (if applied to data) or search time (if to queries), since only two diagonal corner points are needed for each MBR instead of all individual points in it. Thus, in the traditional approach, a low-dimensional MBR is constructed by dividing data or query sequences into window sequences, transforming each (high-dimensional) window sequence to a low-dimensional sequence, and bounding the low-dimensional sequence points into MBRs. This approach requires as many lower-dimensional transformations as the number of window sequences, which can be very large.

I.LITERATURE SURVEY

TRANSFORMATION TECHNIQUE TO BE SAFE[1]

This approach significantly reduces the number of lower-dimensional transformations needed in similar sequence matching. However, it poses a risk that some transformed sequences may fall outside the transformed low-dimensional MBR. propose safe MBR-transformation which has the property that every possible transformed sequence is inside a safe MBR-transformed MBR. Then, considering the discrete Fourier transform (DFT) and the discrete Cosine transform (DCT), we prove that they are not safe as MBR-transformations, and modify them to become safe MBR-transformations (called mbrDFT if DFT-based and mbrDCT if DCT-based). Then, we prove the safeness and optimality of mbrDFT and mbrDCT. Analyses and experiments show that the mbrDFT and mbrDCT reduce the execution time by several orders of magnitude due to the reduction in the number of lower-dimensional transformations. The proposed safe MBR-transformation provides a useful framework for a variety of applications that require a direct transformation of a high-dimensional MBR to a low-dimensional MBR.

PREPROCESSING OF TIME SERIES DATA[2]

A time-series is a sequence of real values sampled at continuous time points. Examples include stock prices, product sales records, medical measurements, and scientific experiments data. For efficient processing of similar time-series matching, much research has been performed over many years. Given a query sequence Q, similar time-series matching algorithms find the time-series whose distances from Q are within the specified threshold. These algorithms assume that all the time-series are stored in the database and are intrinsically static. Distance (similarity)measures that are most widely accepted by the existing algorithms are the Euclidean distance and the dynamic time warping (DTW) distance. However, Keogh identified four kinds of distortions that may produce inaccurate search results: offset translation, amplitude scaling, linear trends, and noise. To remove these distortions, preprocessing is performed on the time-series before they are compared. The preprocessing helps find similar time-series more accurately. An algorithm for distortion-free predictive streaming time-series matching. Distortion-free in the sense that it performs preprocessing on time-series to remove the offset translation and amplitude scaling distortions at the same time. The preprocessing to remove these two distortions is known as normalization transform, and is useful for finding time-series that have similar fluctuation trends. Algorithm predictive in the sense that it performs streaming time-series matching against the predicted most recent subsequences in the near future, and thus improves search performance.

MINIMUM-DISTANCE MATCHING-WINDOW PAIR FOR SUBSEQUENCE[5]

Time-series data are of growing importance in many new database applications such as data mining and data warehousing. A time-series is a sequence of real numbers representing values at specific points in time. Typical examples of time-series data include music data, stock prices and network traffic data. The time series data stored in a database are called data sequences. To perform ranked subsequence matching, the window construction method to be work exploited, Dual Match, which is both efficient and simple for range subsequence matching. That is, divide data sequences into disjoint windows and query sequences into sliding windows. A concept of the minimum-distance matching-window pair and derive a lower bound,called the mdmwp-distance, based on this concept. The minimum-distance matching-window pair for a subsequence is obtained, thus it can derive the mdmwp distance for the subsequence. Then, aggressively prune is done unnecessary subsequence access requests in the index level. It also validates the algorithm by proving the lower-boundness of the mdmwp-distance. A novel optimization technique called deferred group subsequence retrieval to avoid excessive random disk I/Os and bad buffer utilization. Deferred group subsequence retrieval1) delays a set of subsequence retrieval requests, 2) groups the requests by their corresponding subsequences, and 3) enables batch retrieval. Since we have accumulated many requests, it can access subsequences in a sequential fashion. In addition, by exploiting many delayed matching windows, another lower bound,called the window-group distance, that can be used together with deferred group subsequence retrieval. Based on the window-group distance, it perform another ranked subsequence matching algorithm. It also validate this algorithm by proving the lower bound property of the window-group distance.

DTW APPROACH OF TIME SERIES DATA[3]

Time series data is important for commerce, science, and engineering. It frequently serves as a basis for decision and policy-making. Large amounts of time-dependent data are created, acquired, and analysed. Examples include stock market data analysis on prices of publicly traded securities and sensor-based monitoring of seismic activities. A novel approach called Anticipatory Pruning (AP) is introduced for speeding up DTW-based retrieval on both univariate and multivariate time series. Intuitively speaking, AP anticipates the DTW refinement result by exploiting previously computed information in multistep filter- and-refine algorithms. Anticipatory pruning is lossless for a certain class of lower bounding filters.It proves that the most widely used state-of-the-art approaches are in this class. Intuitively, the filters have to be lower bounding themselves, piecewise, and reversible to be usable in this approach. As anticipatory pruning is orthogonal to existing lower- bounding filtering, indexing, and dimensionality reduction techniques, it can be exibly combined with such techniques for additional speed-up. By computing an estimated overall DTW distance from already available filter information, a series of lower bounds of the DTW is derived that requires hardly any overhead. Experimental evaluation demonstrates a substantial reduction in the number of calculations and consequently a significantly reduced runtime. Through this technique can be exibly combined with existing and future DTW lower bounds.

TIME SERIES DATA IN SUBSEQUENCE MATCHING[4]

Time series data naturally appear in a wide variety of domains, including scientific measurements, financial data, and sensor networks, audio, video, and human activity. Subsequence matching is the problem of identifying, given a query time series and a database of time series, the database subsequence (i.e., some part of some time series in the database) that is the most similar to the query sequence. Achieving efficient subsequence matching is an important problem in domains where the database sequences are much longer than the queries, and where the best subsequence match for a query can start and end at any position of any database sequence. Improved algorithms for subsequence matching can make a big difference in real-world applications such as query by humming, word spotting in handwritten documents, and content-based retrieval in large video databases and motion capture databases. EBSM(shorthand for Embedding-Based Subsequence Matching) is presented which is a general method for speeding up subsequence matching in time series databases. The first to explore the usage of embeddings for subsequence matching for unconstrained DTW. The key differentiating features of our method are the following:

EBSM converts, at least partially, subsequence matching under DTW into a much easier vector matching problem. Vector matching is used to identify very fast a relatively small number of candidate matches. The computationally expensive DTW algorithm is only applied to evaluate those candidate matches.

EBSM is the first indexing method, in the context of subsequence matching, that focuses on unconstrained DTW, where optimal matches do not have to have the same length as the query. The only alternative method for this setting, PDTW, which uses piecewise aggregate approximation (PAA), is a generic method for speeding up DTW.

Implementation of PDTW(for the purpose of comparing it to EBSM) is also a contribution, as it differs from the way PDTW has been described by its creators: Add a refine step that significantly boosts the accuracy vs efficiency trade-offs achieved by PDTW.

EBSM provides the best performance in terms of accuracy versus efficiency, compared to the current state-of-the-art methods for subsequence matching under unconstrained DTW: the exact

SPRING method that uses the standard DTW algorithm, and the approximate PDTW method.

III. Proposed model

A new hybrid DCT-SVD similar pattern matching technique is perform in proposed system. Discrete cosine transform (DCT) is widely used in video coding due to its high energy compaction and efficient computation complexity, singular value decomposition (SVD) is a transform that provides optimal energy compaction for any data. DCT and SVD are combined to achieve optimal performance of the transform part. SVD is used only for the blocks for which DCT cannot provide good compression. The decision criterion is set in the DCT domain. By dropping a certain number of coefficients in the DCT domain, the energy loss is calculated. Whether or not sending the block to SVD domain is based on the energy loss.It will take advantages of DCT first, and use SVD only for the blocks that DCT does not compact energy well.Advantages of proposed system, It can improved accuracy in matching more similar pattern and Optimal balancing of energy values.

IV.CONCLUSION

In this paper, we propose a system that provides the better result than the existing approach in terms of improved accuracy.

References

[1]Y.-S. Moon and B.-S. Lee, “Safe MBR-Transformation in SimilarSequence Matching,” Information Sciences, Vol. 270, pp. 28-40, Jun.2014.

[2]W.-K. Loh, Y.-S.Moon, J. Srivastava, Distortion-free predictive streaming time-series matching, Inf. Sci. 180 (8) (2010) 1458–1476.

[3] I. Assent, M. Wichterich, R. Krieger, H. Kremer, T. Seidl, Anticipatory DTW for efficient similarity search in time series databases, in: Proc. of the 35th Int’l Conf. on Very Large Data Bases, Lyon, France, August 2009, pp. 826–837

[4] V. Athitsos, P. Papapetrou, M. Potamias, G. Kollios, D. Gunopulos, Approximate embedding-based subsequence matching of time series, in: Proc. of the ACM SIGMOD Int’l Conf. on Management of Data, Vancouver, Canada, June 2008, pp. 365–378.

[5]W.-S. Han, J. Lee, Y.-S. Moon, H. Jiang, Ranked subsequence matching in time-series databases, in: Proc. of the 33rd Int’l Conf. on Very Large Data Bases, Vienna, Austria, September 2007, pp. 423–434