ISSN 2229-5518

Hybridized Level Transformation Technique for Related Time Series Data

Rosario Arockia Raj R, Devipriya K

ABSTRACT

Sequence pattern matching technique is to find that related time series data but the data has high dimensionality, so it is critical. To improve the accuracy of the sequence pattern matching. DCT and DFT based MBR transformation is introduced to construct a low dimensional MBR which would convert the high dimensional MBR into low dimensional MBR. 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 a 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 DCT-SVD approach provides better result than the existing approach in terms of improved accuracy and performance evaluation.

Key Terms: - Sequence Pattern Matching, Safe MBR Transformation, DCT-SVD Technique, Improved Accuracy, TimeSeriesData.

1 INTRODUCTION

1.1 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.

1.2 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, there by mitigating the need to operate in frequency domain.

1.3 Exploratory analysis

The clearest way to examine to 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 present change per year in this year was calculated.

• *Rosario Arockia Raj R is currently pursuing master's degree program in computer science engineering in Sri Krishna College of engineering and technology Coimbatore, PH-9500241140. E-mail: arockiarose62@skcet.ac.in*
• *Devipriya K is currently working as Assistant Professor in computer science engineering in sri Krishna college of engineering and technology Coimbatore, Email devipriyak@skcet.ac.in*

ISSN 2229-5518

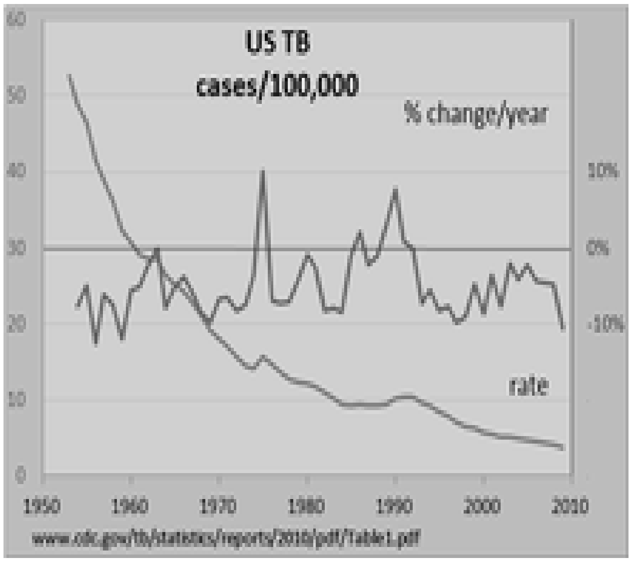

Fig 1.1. Exploratory analysis

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 behavior which need not be related to seasonality. For example, sun spot activity varies over 11 year cycles. Other common examples include celestial phenomena, weather 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.

1.4 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.

2 LITERATURE SURVEY

2.1 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.

2.2 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, that are most widely accepted by the existing algorithms are the Euclidean distance and the dynamic time warping (DTW) distance. The preprocessing helps find similar time-series more accurately. 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.

2.3 MINIMUM-DISTANCE MATCHING-WINDOW PAIR FOR SUBSEQUENCE [11]

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. 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, the minimum-distance matching-window pair for a subsequence is obtained, and thus it can derive the mdmwp distance for the subsequence. A novel optimization technique called deferred group subsequence retrieval to avoid excessive random disk I/O sand bad buffer utilization.

2.4 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 analyzed. A novel approach called Anticipatory Pruning (AP),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.

3 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.

4 MODULE DESCRIPTIONS

4.1 Preprocessing

In statistics, a moving average (rolling average or running average) is a calculation to analyze data points by creating a series of averages of different subsets of the full data set. It is also called a moving mean (MM) or rolling mean and is a type of finite impulse response filter. Variations include: simple, and cumulative, or weighted forms. Simple moving average (SMA) is the unweight mean of the previous then the

$$SMA = \frac{p_M + p_{M-1} + \cdots + p_{M-(n-1)}}{n}$$

4.2 Low dimensional MBR construction

The proposed technique which uses this safe MBR-transformation can drastically reduce the number of lower-dimensional transformations, compared with using the traditional technique which constructs an MBR after tens or thousands of lower-dimensional transformations for individual sequences. Algorithm (called LMBR-mbr) Algorithm bounds high-dimensional sequences into a high-dimensional MBR, one MBR for each r sequences, and transforms each of the resulting MBRs to a low dimensional MBR. This requires only two transformations for each MBR (one for L and one for U of the MBR [L, U]).

4.3 Hybridized DCT-SVD

Discrete cosine transform (DCT) is widely used in 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.

4.4 Sequence matching technique

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" (i.e. straight-line) distance between two points in Euclidean space. A generalized term for the Euclidean norm is the L^2 norm or L^2distance.TheEuclidean distance between points p and q is the length of the line segment connecting them ($\overline{\mathbf{pq}}$).

In Cartesian coordinates, if $p = (p_1, p_2, ..., p_n)$ and $q = (q_1, q_2, ..., q_n)$ are two points in Euclidean n-space, then the distance (d) from p to q, or from qtop is given by

4.5 Performance evaluation

The performance evaluation was conducted to prove the improvement of the proposed research scenario than the existing work in terms of improved accuracy and the time complexity.

5 CONCLUSION

In this paper, sequence pattern technique 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 Similar Sequence 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] Y.-S. Moon, K.-Y. Whang, W.-K. Loh, Duality-based subsequence matching in time-series databases, in: Proc. of the 17th IEEE Int'l Conf. on Data Engineering, Heidelberg, Germany, April 2001, pp. 263–272.

[6] Y.-S. Moon, K.-Y. Whang, W.-S. Han, General match: a subsequence matching method in time-series databases based on generalized windows, in: Proc.of the ACM SIGMOD Int'l Conf. on Management of Data, Madison, Wisconsin, June 2002, pp. 382–393.

[7] Y.-S. Moon, An MBR-safe transform for high-dimensional MBRs in similar sequence matching, in: Proc. of the 12th Int'l Conf. on Database Systems forAdvanced Applications, Bangkok, Thailand, April 2007, pp. 79–90.

[8]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

[9] F. Korn, H.V. Jagadish, C. Faloutsos, Efficiently supporting ad hoc queries in large datasets of time sequences, in: Proc. of the ACM SIGMOD Int'l Conf. on Management of Data, Tucson, Arizona, June 1997, pp. 289–300.

[10] G.R. Hjaltason, H. Samet, Incremental similarity search in multimedia databases, in: Dept. of Computer Science, University of Maryland, College Park, Technical Report 4199, and November 2000.

[11] M.J. Hsieh, M.S. Chen, P.S. Yu, Integrating DCT and DWT for approximating cube streams, in: Proc. of the 14th ACM Int'l Conf. on Information and Knowledge Management, Bremen, Germany, October 2005, pp. 179–186.

[15]K.W. Chu, M.H. Wong, Fast time-series searching with scaling and shifting, in: Proc. of the 18th ACM SIGACT SIGMOD-SIGART Symp. on Principles of Database Systems, Philadelphia, Pennsylvania, June 1999, pp. 237–248