Asian Journal of
Applied
Sciences

A Survey on Effective Pattern Matching in Uncertain Time Series Stream Data

D. Rajalakshmi and K. Dinakaran
R.M.D. Engineering College, Kavaraipettai, Tamil Nadu, India

Corresponding Author: D. Rajalakshmi, R.M.D. Engineering College, Kavaraipettai, Tamil Nadu, India

ABSTRACT

In real time application such as weather forecasting, coal mine surveillance and privacy preserving data streams arrive at a rate higher than in traditional sensing applications. The processing of these raw data must be as fast as stream speed. These applications convert the raw data into a specified pattern and these patterns vary from application to application. It is then subjected to sophisticated query processing to extract high level information. Uncertainty in stream time series may occur for two reasons such as the inherent imprecision of sensor readings or privacy preserving conversion. Representation of uncertain data over stream time-series, uncertain data management and designing of a data mining algorithm on considering the uncertainty affects the data mining process. The processing of data in the real time application is difficult as it is naturally incomplete and noisy and the observed data pattern is different from the actual pattern required for further processing. A major challenge in processing the stream time-series data with uncertainty is to capture uncertainty as data propagates through query operators until the final result and to process the data at stream speed. The modeling of uncertain time-series without affecting the precision of the system still remains a difficult task. More importantly, it should avoid increased false positives. In this study, a survey of time series data management for pattern matching is provided. The similarity search over uncertain data will be explored. The issues in existing research, limitations and methodology for pattern matching on uncertain time series stream data are examined in this study.

Key words: Time series data, dimensionality reduction, similarity measures, stream time series, similarity search

INTRODUCTION

Time series has become increasingly important and common data type for a large number of applications. Therefore, there has been much research effort taken to time series data management in the recent years. The major application of stream time-series data includes internet traffic analysis, sensor network monitoring, moving object search and financial data analysis that requires constant monitoring of stream time series. The frequent updation of data in stream time series and impractical storing of all the data in memory are the unique characterizations of the stream time-series data.

As the network, especially the sensor network widens its application in day-to-day life; one can observe the uncertain data. Uncertain data is incomplete, inaccurate and even erroneous in nature originating from a variety of environments. Uncertainty in time series may occur for various causes

in real-time applications, especially in the data imprecision of sensor observation and privacy-preserving transformation. In the coal mine supervision (Xue *et al.*, 2006) the sensors are deployed throughout the site to gather time-series data for gas density and temperature. For safety reasons, the monitoring system must detect dangers like leakage of gas and fire alarm well in advance. These dangers generally correspond to a specific pattern in a contour-map. It is necessary to check whether those stream time-series matches with any of the predefined patterns obtained from the historic events and alert the miners if any pattern match is detected.

Personal information submitted by individuals and corporations steadily increases and recently developed applications are based on the mining of these data-sets including location-based services and social network applications. Privacy preserving is the major challenge in those applications and can be achieved to some extent using various privacy-preserving transforms (Aggarwal, 2008) such as noisy perturbations, noisy aggregates and reduced granularity. The data can still be queried and retrieved even in the presence of privacy-preserving transforms. It can address the uncertainty introduced by these transforms with some modifications in the existing techniques. The study in Dallachiesa *et al.* (2011) provides a survey of techniques that models and processes uncertain time-series data. The study in Suresh *et al.* (2008) provides an approach, Self Organizing Maps (SOM) to organize gene expression data into clusters so that the function of unknown genes can be discovered.

This study mainly focuses towards pattern matching on uncertain stream time series data that offer better performance with reduced false positives.

There are two main reasons for time-series data to be treated as uncertain. First, physical data collection methods are imperfect. For instance, the accuracy of data collected by the sensor is usually correlated with a certain error distribution. Second, a certain degree of uncertainty is purposely added to time-series so as to preserve privacy. The study of uncertain time-series is important as its existence is widespread and similarity matching serves as the basics for designing a variety of mining algorithms. The study in Aggarwal and Yu (2009) provides a detailed study of uncertain data algorithms. If the uncertain data streams as part of time-series are fed into existing stream processing systems, the results obtained in the end-applications are of poor quality. If any of the monitoring systems ignore the time-series with uncertain values it cannot offer an accurate results in terms of quality and efficiency. The impact of uncertainty results in undesirable and intolerable working conditions in real-time mission critical applications.

In summary, we make the following contributions:

C Review the representation of time series with reduced dimensionality and similarity measures in time series data
C Perform an extensive evaluation for similarity search over time series and uncertain data
C Review the issues in the existing research work on uncertain and time series data
C Propose a methodology for the design and development of a pattern matching system of the data stream that captures data uncertainty

REPRESENTATION OF TIME-SERIES WITH REDUCED DIMENSIONALITY

Generally, the dimensionality of the time-series is high and hence, similarity search over high-dimensional index emerges as a serious issue. The processing of the query pattern over indexes degrades tremendously with the increasing dimensionality. In order to handle the vast dimensionality, a variety of dimensionality reduction methods have been suggested that is to be

executed before indexing. The commonly used dimensionality reduction methods are Singular Value Decomposition (SVD) (Korn *et al.*, 1997), Discrete Fourier Transform (DFT) (Agrawal *et al.*, 1993), Discrete Wavelet Transform (DWT) (Chan and Fu, 1999), Piecewise Aggregate Approximation (PAA) (Keogh *et al.*, 2001), Piecewise Linear Approximation (PLA) (Chen *et al.*, 2007), Principal Component Analysis (PCA) (Valarmathie *et al.*, 2009) and Chebyshev Polynomials (CP) (Cai and Ng, 2004).

The main idea of SVD is to maintain as many eigenvectors as the space restrictions permit. The DFT maps the time sequences with the frequency domain. The DFT is an orthogonal transform that extracts features from the sequence. In DFT, the Euclidean distance in the time domain is maintained in the frequency domain. Hence, DFT satisfies the "completeness of feature extraction" principle. The DWT offers only a coarse representation of original time sequence using its preceding coefficients. The DWT is mainly used for multi-resolution representation of signals. The DWT can provide locations in both time and frequency using time-frequency localization property and it's representation can handle more information.

The PAA is mainly designed for arbitrary length queries that allow constant time insertions and deletions. It handles the weighted Euclidean distance measure. The PLA is a dimensionality reduction approach that offers an efficient similarity search without false positives over time-series databases. The CP obtains coefficients of Chebyshev polynomials that act as the reduced data. Moreover, the time series representation can be broadly classified as model-based, data adaptive, non data adaptive and data dictated. All these dimensionality reduction methods follow the lower bound (Faloutsos *et al.*, 1994), so that they considerably reduce the false positive rate.

Clipping is an effective transformation for time series data mining mainly based on similarity in structure (Bagnall *et al.*, 2006). It is a simple transformation that retains much of the fundamental structural information in the original data. It offers better results and it can be applied in combination with a variety of algorithms.

The approach in Lin *et al.* (2007) presents a novel symbolic representation of time series. It offers dimensionality reduction and moreover, it defines the distance measures in terms of symbolic representation in which the lower bound corresponds to the distance measure of the original time-series data. This kind of symbolic representation allows implementing the data mining algorithms effectively and produces identical results as that of the algorithm operations on the original time-series data. The study in Gullo *et al.* (2009) proposed Derivative time-series Segment Approximation (DSA) representation model to support accurate and fast similarity detection in time series. The DSA is able to transform a time series into a feature rich sequence by combining the notions of derivative estimation, segmentation and segment modeling. The approach piecewise vector quantized approximation in Wang and Megalooikonomou (2008) outperforms PCA techniques in clustering and similarity searches. The study in Wang *et al.* (2013) compares 8 different time-series representation techniques and compares their performance over several time series data sets. The approaches in Lian *et al.* (2009) and Sethukkarasi *et al.* (2010) represents time series using Multi Segment Mean (MSM) representation and multi scale segment median approximation representation for stream time-series image data, respectively for fast pattern matching in stream time.

SIMILARITY MEASURES IN TIME-SERIES DATA

Euclidean distance is the most commonly used similarity measure that can be used with any other type of similarity measures (Agrawal *et al.*, 1993; Faloutsos *et al.*, 1994; Chan *et al.*, 2003).

Though the Euclidean distance is the optimal distance measure, it is proved to be unsuccessful in measuring distances of time-series as it suffers from shifting and scaling. Followed by Euclidean distance measure, Dynamic Time Warping (DTW) (Bernad, 1996; Keogh, 2002), Longest Common Subsequence (LCSS) (Boreczky and Rowe, 1996) and edit distance with real penalty (ERP) (Chen and Ng, 2004) have been suggested to address warps in the temporal dimension. The DTW can handle the time-series with different lengths but fails to obey the triangle inequality. An approach in Keogh (2002) makes DTW indexable by adding the time bound envelopes into the time-series. The R-tree index obtained from DTW (Guttman, 1984) is an inefficient similarity measure, especially in query processing on stream time-series. The ERP supports local time shifting and LCSS offers inaccurate similarity measures. An approach in Megalooikonomou *et al.* (2005) represents time-series using multi resolution vector quantized (MVQ) that encodes the time-series using symbols that matches with hierarchical distance function. An approach in Yang and Shahabi (2004) proposed a similarity measure for MTS (Multivariate Time Series) datasets, Eros which is based on principal component analysis. In precision/recall and time efficiency, Eros outperforms other similarity measures such as ED, WSSVD, DTW, Euclidean distance and dynamic time warping.

SIMILARITY SEARCH OVER TIME-SERIES DATA

An approach in Zhu and Shasha (2003a) observes the correlation between any pair of stream time-series within a sliding window in which DFT summarizes the data. Shift Wavelet Tree (SWT) based on DWT that observes the bursts over the stream time-series data (Zhu and Shasha, 2003b). A whole matching (Agrawal *et al.*, 1993) technique matches data sequences of the same length as that of the query sequence. An approach in Keogh (1997) presents a robust similarity search method that supports scalability in the time axis and is robust to noise, offset translation and amplitude scaling. A probabilistic approach in Keogh and Smyth (1997) locates patterns of interest in massive time series data sets that exploit piecewise linear segmentation for representing data and defines the local features using prior information distribution.

An approach called bit parallel matching on streams (BPS) in Saito *et al.* (2007) investigates the issues in complex matching over a single continuous stream of numerical values and it offers a solution based on bit parallel procedure for string matching. Streaming pattern discovery in multiple time series (SPIRIT) (Papadimitriou *et al.*, 2005) is a technique introduced for discovering the patterns of multiple streams in which it incrementally finds correlation without buffering. Visual Query Language (VQL) (Haigh *et al.*, 2004) retrieves the patterns in a ranked list of matches. The Skyline Index in Li *et al.* (2004) can be coupled with the dimensionality reduction techniques to improve the performance of similarity search.

SIMILARITY SEARCH OVER UNCERTAIN DATA

A probabilistic approach in Yeh *et al.* (2009) processes similarity queries over uncertain data streams, especially the time-series streams called PROUD. The PROUD distance is the sum of the differences of the streaming time series in which each random variable represents the uncertainty of the value in the corresponding time stamp.

DUST is a novel distance measure suggested in Sarangi and Murthy (2010) that is not based on the existence of multiple observations. DUST assumes that all the time-series values follow some definite distribution. DUST considers each value of the time series as a continuous random variable that follows a specific probability distribution.

An approach in Zhao *et al.* (2010) investigates the construction of wavelet decompositions of uncertain data. It depends on a priori knowledge for estimating the probability distribution over uncertain time series. The study in Abfalg *et al.* (2009) formalizes the Dynamic Time Warping (DTW) for uncertain time-series. It introduces two methods to measure the distance between any two uncertain time-series called Probabilistic Bounded Range Query (PBRQ) and Probabilistic Ranked Range Query (PRRQ). A probabilistic model for generating uncertainty for the raw data is suggested in Diao *et al.* (2009). It exploits statistical techniques for capturing the uncertainty pattern as the data propagates through query operators. Moreover, it exploits advanced approximation techniques to tackle with high-volume streams.

Clustering of uncertain data streams is difficult as the uncertainty in attribute values can greatly affect the cluster formation using the data points. An approach in Aggarwal and Yu (2008) exploits a variety of summary structures to track the pattern of the stream in real time and use it for clustering process. An approach in Cormode and McGregor (2008) investigates the issues related to the clustering of uncertain data and suggests some appropriate generalizations of standard clustering optimization principle.

ISSUES IN THE EXISTING RESEARCH WORK

Most of the time-series representation methods do not reduce the dimensionality of the original data. The distance measures defined in the symbolic representation method have little correlation with that of in the original time-series. Among the dimensionality reduction techniques mentioned above, only DWT and DFT deals with the stream time-series. The existing similarity measures do not support online monitoring applications as these are designed for full sequence matching. The existing pattern matching algorithms are susceptible to noise, offset translation and amplitude scaling with various degrees.

The existing approaches for modeling uncertain time-series are based on only two different approaches. The first approach estimates the probability density function over uncertain values based on the prior knowledge. The second approach summarizes the distribution of uncertain data as a result of frequent measurements. These approaches offer results with increased false positives. The real challenge is achieving tradeoff between accuracy and efficiency in uncertain time series data streams.

Limitations: The previous works on pattern matching on archived time series are not suitable for stream time series as they cannot deal with the frequent updates. Concerned with the data streams, the efficiency of the computation process is the key issue as the data can be processed only once during the entire computation in the presence of uncertain data. Generally, processing of raw uncertain data is difficult because of its mass volume and stringent time schedule. Raw uncertain data generally comprise of noise that interrupts the pattern matching process and produces false positives. All the time series dependent applications require immediate responses and cannot prefer any post processing. Striking solution between the accuracy and efficiency is highly challenging in the presence of uncertainty.

METHODOLOGY

Figure 1 shows the design and development of a pattern matching system of the data stream that captures data uncertainty from data collection to query processing to obtain final pattern that matches the query pattern. The incoming patterns are initially transformed into streams with uncertain data to model the core data generation process. The data of interest can be inferred from this model. The inference process computes a distribution of values for the uncertain data required

Asian J. Applied Sci., 8 (3): 217-226, 2015

Table 1: Summarization of existing work

| Titles	Authors	Algorithm	Advantage	Disadvantage	Metrics
Efficient similarity search in sequence databases (Agrawal et al., 1993)	R. Agrawal, C. Faloutsos and A.N. Swami	DFT: Mapping R-trees: Indexing	Avoids the dimensionality curse	Introduces few false hits	Total execution time
Fast subsequence matching in time-series databases (Faloutsos et al., 1994)	C. Faloutsos, M. Ranganathan and Y. Manolopoulos	ST-index	Short response time	It has limitations on the shortest possible query and queries greater than a certain length, affects the speed of the matching	Response time
A fast and robust method for pattern matching in time series databases (Keogh, 1997)	Eamonn Keogh	Bottom-up algorithm	Robust to scaling in time axis and noise	Criteria for selecting the pair of segments to merge in each iteration is not specified	Speed-up
Efficiently supporting ad hoc queries in large datasets of time sequences (Korn et al., 1997)	F. Korn, H. Jagadish and C. Faloutsos	Singular Value Decomposition (SVD) SVDD as enhancement	Easy accessible to a huge database	It is not robust to outliers and it suffers from practical issues	Reconstruction error, storage space
Probabilistic similarity search for uncertain time series (Abfalg et al., 2009)	Johannes Abfalg, Hans-Peter Kriegel, Peer Kroger, Matthias Renz	Probabilistic Bounded Range Queries (PBRQ) Probabilistic Ranking Range Query (PRRQ)	Provide quality guarantees	It is applicable only in the cases where the standard deviation of the error is relatively small and the length of the time series is also small	Speed-up factor
PROUD: A probabilistic approach to processing similarity queries over uncertain data streams (Yeh et al., 2009)	M. Yeh, K. Wu, P. Yu and M. Chen	PROUD	Provide quality guarantees	It is difficult to set the value of probabilistic threshold	Miss ratio, false alarm ratio, computation time cost
DUST: A generalized notion of similarity between uncertain time series (Sarangi and Murthy, 2010)	S. Sarangi and K. Murthy	DUST	Provide quality guarantees	Performs better only if the distribution of the observation errors is mixed and the parameters of this distribution are known	Accuracy
On wavelet decomposition of uncertain time series data sets (Zhao et al., 2010)	Y. Zhao, C. Aggarwal and P. Yu	Wavelet decomposition algorithm	Improves the compression rate of the representation	Prior knowledge is required to estimate the probability density function	Accuracy, execution time

222
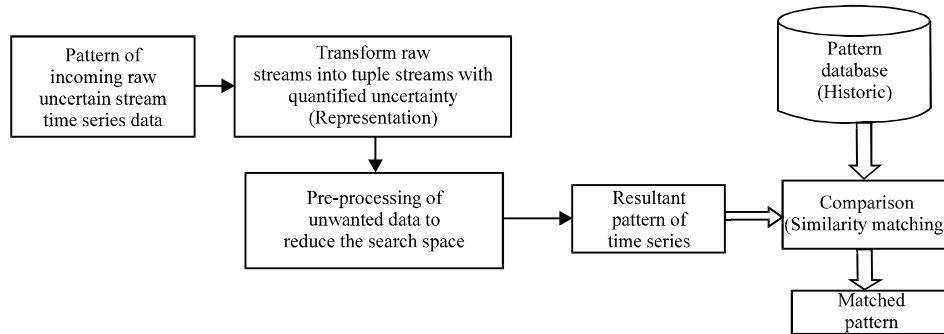
Fig. 1: Pattern matching over uncertain stream time series data

in later processing. The uncertain data are then eliminated to prune off large search space using any of the lower bound distance measures. The size of the search space varies exponentially with the number of features in the query pattern. Therefore, it becomes necessary to depend on effective search techniques. The sensor data are generally updated in a periodic manner. This necessitates the use of indexing system to update the current attribute value. Each part of the query is associated with a range of possible values and probability density function which quantifies the behavior of the data over that range. The obtained resultant pattern is compared with the time-series database to provide the final pattern that matches with the query pattern. Table 1 shows the previous works on time series data.

CONCLUSION

The area of uncertain data management in stream time series application has been treated as an unsolved issue. This work presented an overview of pattern matching methods over uncertain time series stream data. This study also presented the commonly used similarity measures, similarity searches on stream time series and uncertain data along with the representational issues in uncertain data management. The accuracy of the pattern matching algorithm is measured in terms of precision and recall. If the uncertainty is perfectly mined out, stream time series data can be extended to numerous applications.

REFERENCES

Abfalg, J., H.P. Kriegel, P. Kroger and M. Renz, 2009. Probabilistic similarity search for uncertain time series. Proceedings of the 21st International Conference on Scientific and Statistical Database Management, June 2-4, 2009, New Orleans, LA., USA., pp: 435-443.

Agrawal, R., C. Faloutsos and A.N. Swami, 1993. Efficient similarity search in sequence databases. Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms, October 13-15, 1993, Chicago, IL., USA., pp: 69-84.

Aggarwal, C.C., 2008. On unifying privacy and uncertain data models. Proceedings of the IEEE 24th International Conference on Data Engineering, April 7-12, 2008, Cancun, Mexico, pp: 386-395.

Aggarwal, C.C. and P.S. Yu, 2008. A framework for clustering uncertain data streams. Proceedings of the 24th International Conference on Data Engineering, April 7-12, 2008, Cancun, Mexico, pp: 150-159.

Aggarwal, C.C. and P.S. Yu, 2009. A survey of uncertain data algorithms and applications. IEEE Trans. Knowledge Data Eng., 21: 609-623.

Bagnall, A., E. Keogh, S. Lonardi and G. Janacek, 2006. A bit level representation for time series data mining with shape based similarity. Data Mining Knowledge Discov., 13: 11-40.

Bernad, D.J., 1996. Finding Patterns in Time Series: A Dynamic Programming Approach. In: Advances in Knowledge Discovery and Data Mining, Fayyad, U.M. (Ed.). AAAI Press, Menlo Park, CA., USA., ISBN-13: 9780262560979, pp: 229-235.

Boreczky, J.S. and L.A. Rowe, 1996. Comparison of video shot boundary detection techniques. J. Electron. Imaging, 5: 122-128.

Cai, Y. and R. Ng, 2004. Indexing spatio-temporal trajectories with Chebyshev polynomials. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 13-18, 2004, Paris, France, pp: 599-610.

Chan, K.P. and A.C. Fu, 1999. Efficient time series matching by wavelets. Proceedings of the 15th International Conference on Data Engineering, March 23-26, 1999, Sydney, Australia, pp: 126-133.

Chan, F.P., A.C. Fu and C. Yu, 2003. Haar wavelets for efficient similarity search of time-series: With and without time warping. IEEE Trans. Knowledge Data Eng., 15: 686-705.

Chen, L. and R. Ng, 2004. On the marriage of Lp-norms and edit distance. Proceedings of the 30th International Conference on Very Large Data Bases, Volume 30, August 31-September 3, 2004, Toronto, Canada, pp: 792-803.

Chen, Q., L. Chen, X. Lian, Y. Liu and J.X. Yu, 2007. Indexable PLA for efficient similarity search. Proceedings of the 33rd International Conference on Very Large Data Bases, September 23-27, 2007, University of Vienna, Austria, pp: 435-446.

Cormode, G. and A. McGregor, 2008. Approximation algorithms for clustering uncertain data. Proceedings of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, June 9-12, 2008, Vancouver, Canada, pp: 191-200.

Dallachiesa, M., B. Nushi, K. Mirylenka and T. Palpanas, 2011. Similarity matching for uncertain time series: Analytical and experimental comparison. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Querying and Mining Uncertain Spatio-Temporal Data, November 1-4, 2011, Chicago, IL., USA., pp: 8-15.

Diao, Y., B. Li, A. Liu, L. Peng, C. Sutton, T. Tran and M. Zink, 2009. Capturing data uncertainty in high-volume stream processing. Proceedings of the 4th Biennial Conference on Innovative Data Systems Research, January 4-7, 2009, Asilomar, CA., USA., pp: 1-11.

Faloutsos, C., M. Ranganathan and Y. Manolopoulos, 1994. Fast subsequence matching in time-series databases. Proceedings of the ACM SIGMOD International Conference on Management of Data, May 24-27, 1994, Minneapolis, MN., USA., pp: 419-429.

Gullo, F., G. Ponti, A. Tagarelli and S. Greco, 2009. A time series representation model for accurate and fast similarity detection. Pattern Recognit., 42: 2998-3014.

Guttman, A., 1984. R-Tree: A dynamic index structure for spatial searching. Proceedings of the 13th ACM SIGMOD International Conference on Management of Data, June 18-21, 1984, Boston, USA., pp: 47-57.

Haigh, K.Z., W. Foslien and V. Guralnik, 2004. Visual query language: Finding patterns in and relationships among time series data. Proceedings of the 7th Workshop on Mining Scientific and Engineering Datasets, April 24, 2004, Lake Buena Vista, USA., pp: 1-8.

Keogh, E., 1997. A fast and robust method for pattern matching in time series databases. Proceedings of the SAS Conference on Western Users of SAS Software, October 22-24, 1997, California, USA., pp: 145-150.

Keogh, E.J. and P. Smyth, 1997. A probabilistic approach to fast pattern matching in time series databases. Proceedings of the 3rd International Conference of Knowledge Discovery and Data Mining, August 14-17, 1997, Newport Beach, CA., USA., pp: 24-30.

Keogh, E., K. Chakrabarti, M. Pazzani and S. Mehrotra, 2001. Dimensionality reduction for fast similarity search in large time series databases. Knowledge Inform. Syst., 3: 263-286.

Keogh, E., 2002. Exact indexing of dynamic time warping. Proceedings of the 28th International Conference on Very Large Data Bases, August 22-23, 2002, Hong Kong, China, pp: 406-417.

Korn, F., H.V. Jagadish and C. Faloutsos, 1997. Efficiently supporting ad hoc queries in large datasets of time sequences. ACM SIGMOD Rec., 26: 289-300.

Li, Q., B. Moon and I.F.V. Lopez, 2004. Skyline index for time series data. IEEE Trans. Knowl. Data Eng., 16: 669-684.

Lian, X., L. Chen, J.X. Yu, J. Han and J. Ma, 2009. Multiscale representations for fast pattern matching in stream time series. IEEE Trans. Knowledge Data Eng., 21: 568-581.

Lin, J., E. Keogh, L. Wei and S. Lonardi, 2007. Experiencing SAX: A novel symbolic representation of time series. Data Mining Knowledge Discov., 15: 107-144.

Megalooikonomou, V., Q. Wang, G. Li and C. Faloutsos, 2005. A multiresolution symbolic representation of time series. Proceedings of the 21st International Conference on Data Engineering, April 5-8, 2005, Tokyo, Japan, pp: 668-679.

Papadimitriou, S., J. Sun and C. Faloutsos, 2005. Streaming pattern discovery in multiple time-series. Proceedings of the 31st International Conference on Very Large Data Bases, October 4-6, 2005, Trento, Italy, pp: 697-708.

Saito, T., T. Kida and H. Arimura, 2007. An efficient algorithm for complex pattern matching over continuous data streams based on bit-parallel method. Proceedings of the IEEE International Workshop on Databases for Next Generation Researchers, April 15, 2007, Istanbul, Turkey, pp: 13-18.

Sarangi, S.R. and K. Murthy, 2010. DUST: A generalized notion of similarity between uncertain time series. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 25-28, 2010, Washington, DC., USA., pp: 383-392.

Sethukkarasi, R., D. Rajalakshmi and A. Kannan, 2010. Efficient and fast pattern matching in stream time series image data. Proceedings of the 1st International Conference on Integrated Intelligent Computing, August 5-7, 2010, Bangalore, India, pp: 130-135.

Suresh, R.M., K. Dinakaran and P. Valarmathie, 2008. Clustering gene expression data using self-organizing maps. J. Comput. Applic., 1: 6-7.

Valarmathie, P., M.V. Srinath and K. Dinakaran, 2009. An increased performance of clustering high dimensional data through dimensionality reduction technique. J. Theor. Applied Inform. Technol., 5: 731-733.

Wang, Q. and V. Megalooikonomou, 2008. A dimensionality reduction technique for efficient time series similarity analysis. Inform. Syst., 33: 115-132.

Wang, X., A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann and E. Keogh, 2013. Experimental comparison of representation methods and distance measures for time series data. Data Mining Knowledge Discovery, 26: 275-309.

Xue, W., Q. Luo, L. Chen and Y. Liu, 2006. Contour map matching for event detection in sensor networks. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 27-29, 2006, Chicago, IL., USA., pp: 145-156.

Yang, K. and C. Shahabi, 2004. A PCA-based similarity measure for multivariate time series. Proceedings of the 2nd ACM International Workshop on Multimedia Databases, November 8-13, 2004, Washington, DC., USA., pp: 65-74.

Yeh, M.Y., K.L. Wu, P.S. Yu and M.S. Chen, 2009. PROUD: A probabilistic approach to processing similarity queries over uncertain data streams. Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, March 23-26, 2009, Saint-Petersburg, Russia, pp: 684-695.

Zhao, Y., C. Aggarwal and P. Yu, 2010. On wavelet decomposition of uncertain time series data sets. Proceedings of the 19th ACM International Conference on Information and Knowledge Management, October 26-30, 2010, Toronto, ON., Canada, pp: 129-138.

Zhu, Y. and D. Shasha, 2003a. Efficient elastic burst detection in data streams. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 24-27, 2003, Washington, DC., USA., pp: 336-345.

Zhu, Y. and D. Shasha, 2003b. Warping indexes with envelope transforms for query by humming. Proceedings of the ACM SIGMOD International Conference on Management of Data, June 9-12, 2003, San Diego, CA., USA., pp: 181-192.