Efficient Hemodynamic Event Detection Utilizing Relational Databases and Wavelet Analysis

M Saeed, RG Mark

Harvard - MIT, Cambridge, USA

Abstract

Development of a temporal query framework for time-oriented medical databases has hitherto been a challenging problem. We describe a novel method for the detection of hemodynamic events in multiparameter trends utilizing wavelet coefficients in a MySQL relational database. Storage of the wavelet coefficients allowed for a compact representation of the trends, and provided robust descriptors for the dynamics of the parameter time series. A data model was developed to allow for simplified queries along several dimensions and time scales. Of particular importance, the data model and wavelet framework allowed for queries to be processed with minimal table-join operations. A web-based search engine was developed to allow for user-defined queries. Typical queries required between 0.01 and 0.02 seconds, with at least two orders of magnitude improvement in speed over conventional queries. This powerful and innovative structure will facilitate research on large-scale time-oriented medical databases.

1. Introduction

With advances in computer networking and information storage technologies, the amount of data being produced and stored is growing at a significant rate. In several domains, it is highly desirable to retrieve and analyze data using modern relational database systems and the internet. Often, the data has several time varying components. For example, in finance, one may seek to analyze time series data of stock quotes with respect to various economic indices. Time series analysis may more often have trend queries rather than the conventional point query. The difference between the two classes of queries can be illustrated in the domain of hospital intensive care unit (ICU) patient monitoring. For example, consider the following queries:

Q1: Select all patients that had a measured heart rate greater than 120 beats per minute (bpm).

Q2: Select all patients that had heart rate fluctuations within 8 hours that were greater than 25 bpm.

As a point query, Q1 is comparing all the stored heart rate measurements to a user-specified threshold, and is insensitive to time. Q2 is a trend query and analyzes the time-dependent dynamics of the heart rate trends. Frequently, the *changes* in physiological and clinical variables can offer additional insight into the health of a patient.

Structured Query Language (SQL) can efficiently support point queries, such as Q1. However trend (temporal) queries cannot be efficiently implemented in conventional SQL. Often, numerous "self-join" table operations and logical comparisons are needed such that temporal queries become computationally intractable in large databases. To accommodate temporal queries, the TSQL standard was proposed [1]. However, a shortcoming of TSQL is that it is often necessary to specify time intervals prior to the analysis of the data. Moreover, TSQL is not optimized for multichannel time series data with high sampling rates.

Researchers have also proposed several methods of pre-processing time series data prior to storage in a database. These methods include curve fitting and approximation, moving-averages, singular value decomposition (SVD), and Fourier analysis [2]. In addition to their computational complexity, a significant drawback of the aforementioned techniques is that these algorithms require an *apriori* assumption about the interval length of the temporal query. Shahabi et al. introduced a wavelet data structure for temporal queries and applied the method to financial time series and synthetic data [2]. The algorithm was based on a dyadic wavelet expansion of time series data and required reconstruction of the original time series upon a user-defined query. In addition to the increased storage overhead of their method, the algorithm was not implemented in a relational database.

In this paper, we introduce an overcomplete wavelet structure for temporal queries. We demonstrate that our algorithm is efficient and can be implemented in a modern relational database. We have developed a web-based search engine to support temporal queries of a large ICU patient database. We include examples of user-defined queries and a visualization of the retrieved events from all patient physiological time series.

2. Wavelet-based temporal queries

In this section, we explain the design of our novel wavelet data structure for temporal queries. First, we include a brief set of definitions. A "patient-record" includes all the time series information (channels) for a particular patient. For each time series of each patient record, we compute an overcomplete wavelet transform as described in [3]. An "event" is a location in time of a patient record that satisfies the conditions of a user-specified query. A "query" is a user-defined request for the periods of time in a database in which a logical combination of channels of a record deviate from their respective local mean values by more than the user-specified thresholds. A time "scale" is the approximate interval of time over which an event occurs.

Wavelet analysis can be described as the representation/decomposition of signals with simpler functions at specific scales and locations in time. For example, in Figure 1, a time series, X(t) has a local change of 2H. The wavelet coefficient, w_{jk} (where j=scale, and k=shift in time from the origin), at this location and scale can be calculated by the dot product of W(t) and X(t), and in this example will equal H/2 (assuming the increase occurs over a period of t=2). Dyadic dilations and shifts of W(t) can be used to analyze the changes of X(t) at different locations and scales. In general, multiplication of the wavelet coefficient by a factor of 4 will yield the *approximate* change of X(t) at the respective scale and time. See [4] for a more complete review of wavelet analysis and applications.

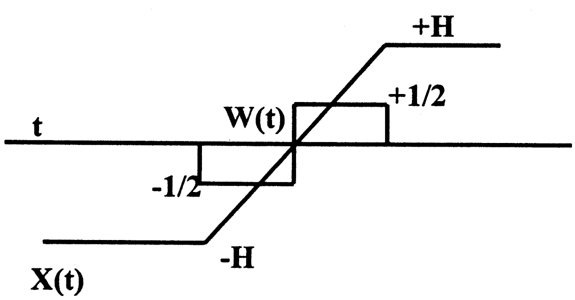

Figure 1. Wavelet approximation of change in X(t)

In the conventional discrete wavelet transform (DWT), wavelets of the same scale are shifted so that there is no overlap between adjacent wavelets. We use the overcomplete wavelet expansion in which the shift is smaller (half a period instead of a full period) so that adjacent wavelets do overlap. This overlap allows for the detection of events that span between two wavelets at the same scale.

Figure 2 includes a flow diagram that describes our overall framework.

We evaluated our temporal query method using data from 57 patients in the MIMIC (Multiparameter Intelligent Monitoring for Intensive Care) in our database [5]. For each patient, we analyzed the heart rate, mean arterial blood pressure, diastolic pulmonary artery pressure, and estimated cardiac output trends as described in [3] using the Haar wavelet family. Additionally, we constructed time series for a subset of continuous medications administered to each patient that may impact the hemodynamics and cardiac function of a patient. We also constructed a time series for net fluid balance by the simple integration of urine output and I.V. input. Insensitive fluid losses were not modeled, but can be added in future implementations.

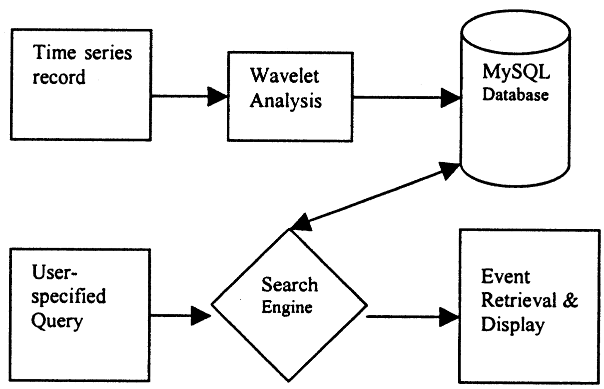

Figure 2: Flow diagram of hemodynamic event search engine.

We created a data model using a MySQL database. The data table for the physiological parameters included the local mean values for each parameter as well as the respective wavelet coefficients and one time stamp. To achieve more compact tables and reduce storage and memory requirements, we applied simple thresholding to prune our data structure. If the wavelet coefficients at a particular location in time had values approximately equal to zero, they were discarded and not stored in the database.

Patient monitoring data is inherently noisy. To reduce the number of "false-positive" events that are retrieved by the search engine, we applied the methodology as described in our previous work [3] to detect and discard intervals of time that were deemed noisy.

To allow for user-specified queries, we developed a web search engine using HTML and the PHP scripting language interfaced to an APACHE web server and the MySQL database. Using conventional HTML forms, the user can specify a set of physiological parameters for analysis using temporal queries. The user can specify the length of the time intervals (scales) for the temporal query, as well as the logical (and/or) relationships between changes (magnitude and direction) in different parameters. The user can also specify changes in medications and/or fluid balance to further refine an event.

Based upon the user-defined query, the search engine will return a summary of the entire database in which the number of "events" that satisfy the query will be noted for each patient. The user can then specify a specific record to be displayed. Using MATLAB, an image is dynamically created with all the trends of the specified patient, and the events were demarcated by the search engine. The image is retrieved by the web-browser for near real-time visualization of the events satisfying the temporal query. The user can subsequently review other patient records with such events.

3. Experimental results

Our wavelet data structure supported efficient retrieval of hemodynamic events from the MIMIC database upon a user-specified query. As can be seen from Table 1, the wavelet-based algorithm resulted in at least two orders of magnitude improvement in retrieval speed for temporal queries. The results of Table 1 are based on 5 typical queries with two or more parameters at one hour and four-hour time scales.

Wavelet-based algorithm	**Ad-Hoc Conventional approach**
0.01-0.02 seconds	>2.6 seconds

Table 1: Comparison of retrieval speed of wavelet-based algorithm and conventional SQL approach

Figure 3 includes a visualization of the retrieved events from a user-specified query. The query sought all events defined by an increase in heart rate over one hour by more than 10 bpm and simultaneous increase in mean arterial blood pressure by more than 10 mmHg during a period of time when the medication, Dopamine, was increased. The three events that were identified are correct as validated by human expert inspection.

4. Conclusions

In this paper, we have described a novel wavelet-based architecture for generalized temporal queries. We have applied our framework to the MIMIC database in the domain of hemodynamic event detection. Using our search engine, a user is able to specify an event of interest as a temporal query. Our framework allows for typical queries to be processed within 0.01-0.02 seconds with at least a two-orders of magnitude improvement in retrieval speeds over conventional techniques.

The search engine was not rigorously evaluated for sensitivity and specificity. In future evaluations of the performance of the search engine, there will be several issues to consider. A standard needs to be developed for defining precisely what a "change" in a time series is. Once a standard is developed, and human experts annotate a set of records, it will be possible to evaluate the performance of our search engine. The search engine performance needs to be evaluated for several different scales and user-specified thresholds.

As we are developing larger-scale temporal medical databases (thousands of patients), our search engine will prove useful in organizing, and searching through such a database. In examining the complex event described in Figure 3, it would be cumbersome to manually find such events in a massive temporal database.

There are several applications of our algorithm in other domains besides ICU patient monitoring. For example, analysis of the fluctuations in the weights (and other parameters) of patients with congestive heart failure, or blood sugar levels in diabetics, can be readily accomplished using our data structure and search engine.

We are currently investigating more intelligent methods of retrieving and ranking the importance of events. In some cases, it is desirable to retrieve only the most important events so that the user is not overwhelmed with data. In order to minimize the storage and memory requirements of our data structure, we applied simple thresholding to the wavelet coefficients and discarded those coefficients with insignificant energy. We plan to investigate more sophisticated techniques in addition to simple thresholding.

Acknowledgements

The National Aeronautics and Space Administration supported this work through the NASA Cooperative Agreement NCC 9-58 with the National Space Biomedical Research Institute. Mr. Saeed was supported through a fellowship from Philips Medical Systems.

References

[1] Soo MD, Jensen CS, Snodgrass. An algebra for TSQL2. In The TSQL2 Temporal Query Language, pages 505-546.Kluwer Academic Press, September 1995.

[2] Shahabi CS, Chung S, Safar M. 2D TSA-tree: A Wavelet-Based Approach to Improve the Efficiency of Multilevel Surprise Mining. PAKDD International Workshop on Mining Spatial and Temporal Data, 2001.

[3] Saeed M. Mark RG, Multiparameter Trend Display using Wavelet Analysis. Proceedings of IEEE Computers in Cardiology, 2000.

[4] Mallat S, A Theory for Multiresolution Signal Decomposotion; the Wavelet Representation. IEEE Trans. On Pattern and Machine Intelligence. 11(7):674-693, July

1989.
[5] Moody GB, Mark RG, Goldberger AL. Physionet: A web-based resource for the study physiologic signals. IEEE Eng. Med. and Bio. 2001;1:70-5.

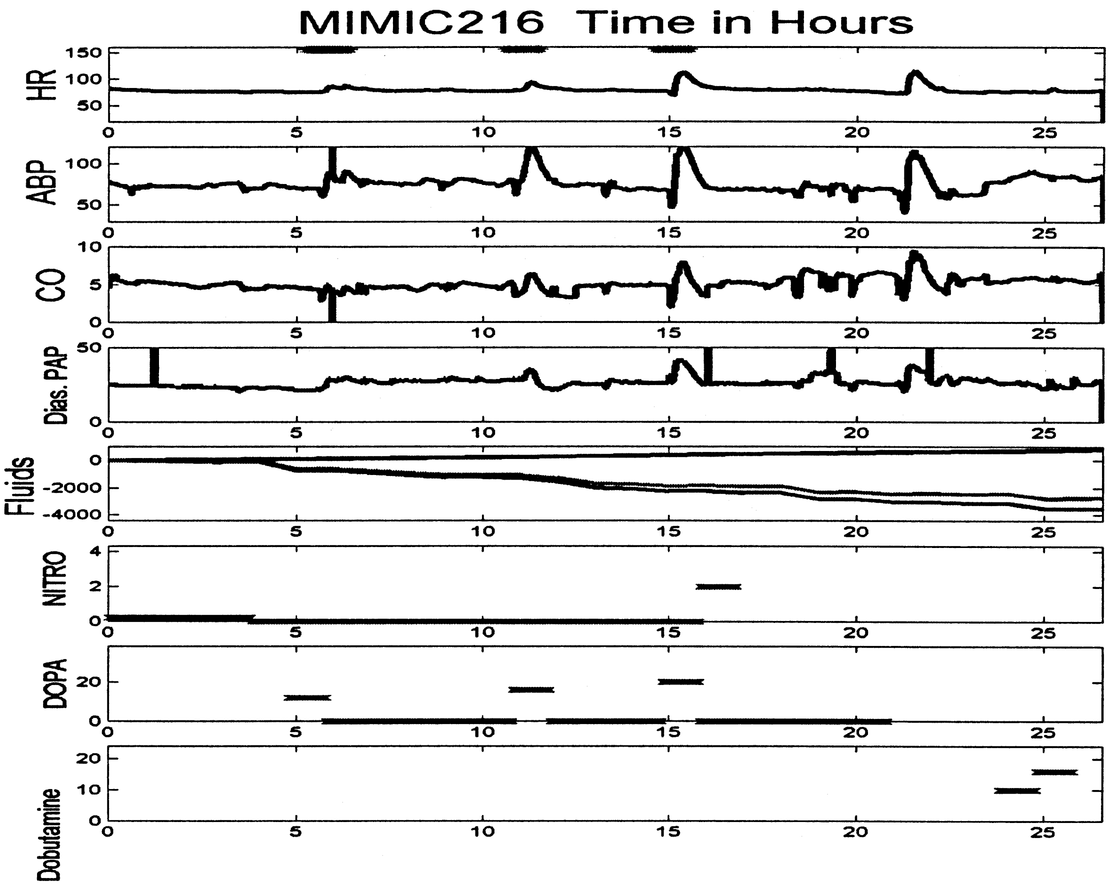

Figure 3: Events retrieved from MIMIC216 based upon a user-specified query for simultaneous increases in heart rate and mean arterial blood pressure during an increase in the administratrion of the vasoconstrictor, Dopamine. The events are demarcated with bars over the heart rate time series at approximately hours 5, 11, and 15.

Address for correspondence.

Mohammed Saeed
E25-505
MIT
77 Massachusetts Avenue
Cambridge, MA, USA 02139
Email: msaeed@mit.edu