An Infrastructure of Stream Data Mining, Fusion and Management for Monitored Patients

Hyoil Han[1], Han C. Ryoo[2], and Herbert Patrick[3]
College of Information Science and Technology, Drexel University[1]
School of Biomedical Engineering, Science and Health Systems[2]
College of Medicine, Drexel University[3]
3141 Chestnut Street, Philadelphia, PA 19104, USA
{hyoil.han, hanryoo}@drexel.edu, Herbert.Patrick@DrexelMed.edu

Abstract

This paper proposes an infrastructure for data mining, fusion and patient care management using continuous stream data monitored from critically ill patients. Stream data mining, fusion, and management provide efficient ways to increase data utilization and to support knowledge discovery, which can be utilized in many clinical areas to improve the quality of patient care services. The primary goal of our work is to establish a customized infrastructure model designed for critical care services at hospitals. However this structure can be easily expanded to other areas of clinical specialties.

1. Introduction

Under current clinical environments, the physiological state of critically ill patients are continuously monitored by various patient monitoring systems that produce multi-channel data streams in time. These signals may contain important physiologic features to explain impending threats to patients' lives. Due to limitations in existing patient monitors and lack of networking infrastructure in hospitals, however, these data streams are generally discarded by the system itself without being stored for data sharing and later used for further analysis and research. Presently, there are no published reports on data archives of medical stream data or on the design of databases for continuous query systems for patient monitoring/diagnostic purposes.

Traditional databases store snapshot data unless it has temporal attributes, such as valid time and transaction time. While a traditional database model works fine with offline and discrete business data, the data from emerging applications requires online data analysis and continuous query processing and thus do not fit into a traditional database model. Monitored patient data are continuously acquired along with time stamps and physiological changes that occur in sequence. These characteristics require query processing to take into account data order and time. This makes stream data processing different from query processing in traditional database management systems (DBMS). Stream data is continuously flowing and the use of it requires an approximate summary, such as a synopsis and thensynopses querying, rather than exact query processing as in DBMS.

Our ultimate goal is to design and build a system model that can alert medical staff in real time to critical changes in a patient's physiology so that clinicians can initiate medical intervention earlier to save lives. The goal of this paper is to propose an infrastructure for advanced patient care systems that incorporates many subsystems for stream data mining, clinical information management, knowledge sharing and semantic interoperability

The objective toward the ultimate goal is to build a prototype which employs data mining techniques to extract medical knowledge and to fuse information obtained from data streams from monitored patients in intensive care units (ICU) with symptom-specific diseases. The central hypothesis is that online and offline data mining and management will be able to predict and detect anomalous medical symptoms much earlier than the medical staff with at least the same or better accuracy.

This paper introduces a new project idea on basic research in combing stream data mining/management with information fusion of monitored patients in an Intensive Care Unit (ICU) setting.

The remainder of this paper is organized as follows: in Section 2, we review related work; in

Section 3 we present the infrastructure of stream data management/mining/fusion for monitored patients. A short conclusion ends the article.

2. Related Work

Emerging applications, such as sensor networks, requires new data models and database management architectures. Data from sensor networks, such as data from bedside monitors in hospitals, creates stream data whose arrival is often in a burst form and whose update frequency is very high. Stream data management is an ongoing research issue in the information and computer science community. Most research into stream data management [1] [2] [3] is related to stream data query processing and stream data management. Work [4] has described a framework for stream data mining with algorithms, but does not describe an infrastructure for stream data management and fusion. Other research [5] describes how to collect body-sensor data for medical information systems and presents database architecture to manage the data. However, this work in [5] does not describe how to mine the collected data to utilize it further for data fusion.

Stream data from sensors is ordered and time stamped. The sensor data warehouse that we will build is a temporal database with historical data that was extracted from stream data. The existing data mining techniques are mainly applied to snapshot data without temporal considerations. Most techniques developed for stream data is related to non-medical applications. Therefore, we cannot apply the existing data/stream mining techniques to our work directly. Temporal data mining is one methodology used to analyze temporal data by transforming time-stamped data into interval-based data. Recent work introduces the abstraction method with two levels of symbolic interpretation for decision support [6]. The work in [7] is extending temporal databases to accommodate rich semantics. The work in [8] describes decision-support systems using probabilistic-network model for data intensive clinical care environements.

3. Tasks and Methods

Our work presents a new infrastructure, which includes not only stream database management but also stream data mining for data fusion. Figure 1 shows an infrastructure, StreDaMiMa, we propose to manage and mine stream data. StreDaMiMa consists of several data engines, such as Archive, Stream Data Engine, Sensor Data Warehouse, Online Data Engine, Offline Data Engine, and User Interface Engine.

3.1 Stream Data Collection

Patients' physiology in ICU are always monitored by patient monitoring systems which display both raw waveforms and computed numeric values of vital signs in various channels on the screen. Such monitors are equipped with invasive, non-invasive or partially invasive sensors cautiously placed on the body surface or pulmonary artery. Typical examples of vital signs are heart rate, respiration rate, body temperature, blood pressure and pulse oximetry.

Data streams of vital signs will be electronically downloaded from such monitors via serial communication ports supported by an interface software and sequentially saved in database for data analysis.

Along with stream data, lab variables such as results of blood test and x-ray readings by clinicians will be accompanied by time stamps of blood drawn and x-ray taken. These lab variables can be incorporated with stream data in terms of time stamps so that we can see correlations in time among various data channels. Clinical diagnostic criteria will be basically used to create medical query systems and to develop diagnostic algorithms to meet medical needs primarily in the ICU. Medical symptoms diagnosed by clinicians will also be recorded from medical charts or records to evaluate the system performance.

3.2 Stream data management and warehousing

The archive in Figure 1 stores raw stream data from bedside monitors and is used to backup all patient sensor data for recovery. The Stream Data Engine consists of the Stream Data Buffer Manager (SDBM) and the Stream Data Extraction Query Manager (SDEQM). Stream data arrives continuously and requires an infinitely large main memory to perform real time responses. However, creating an infinitely large main memory is not possible. The Stream Data Buffer Manager (SDBM) manages the buffer by maintaining stream data queues in a small and fast memory cache without degrading the overall stream data mining/fusion/management performance. The stream Data Extraction Query Manager (SDEQM) creates queries issued to the data stream to determine which stream data portion should be stored in the Sensor Data Warehouse, which consists of a bundle of databases. These queries are adaptive to the input

Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)

IEEE COMPUTER SOCIETY

stream data to maintain a reasonable amount of stored relations in the Sensor Data Warehouse.

To build a Sensor Data Warehouse, queries are defined to describe which data portion from among the streamed data from bedside monitors should be extracted. These queries are adaptive to the environment data, such as the current data stream and the medical rules that were discovered from previous mining results. The query adaptability gives flexibility to the amount of streamed data that is to be extracted, because streamed data are from sensors attached to patients in the ICU and the amount of data is huge. Therefore, it is not possible to store all the sensor data in a sensor database. Using query adaptability, we propose to selectively store streamed data in the Sensor Data Warehouse, which is built in relational databases. This is due to the amount of stream data. For example, when a certain emergency occurs, the extraction query manager modifies the extraction query to extract more data from the input stream data in the same interval. Otherwise, the extraction query extracts from the streamed data based on the frequency define for the given time interval. This time interval is adaptive to the environment data. The Sensor Data Warehouse stores relational data extracted from stream data based on extraction queries. Whereas the sensor database stores historical data, streamed data is a data flow that is time stamped and considered as time-series data.

3.3 Online and Offline data engines sensors

The online Data Engine consists of a Stream Data Mining Module (SDM), a Data Fusion Module (DF), and a Continuous Query Processor (CQP). The Data Mining Module discovers (or learns) rules from time series data in a real-time mode. The learned rules are stored in the Medical RuleBase and are applied to the Fusion Module, which detects the emergency state and other critical conditions. Continuous queries against streamed data are applied continuously over time to the input stream after being issued once, and are processed in a real-time mode over variable windows.

The Offline Data Engine consists of a Temporal Data Query Processor (TDQP) and an Offline Data Mining Module (ODM). Continuous queries issued against the Sensor Data Warehouse are processed in the Temporal Data Query Processor. The Temporal Data Query Processor processes aggregate functions over stored relational data in a given time interval. The Offline Data Mining Module learns rules over the stored relational data. Traditional data-mining techniques mainly work with static snapshot data, whereas sensor data has a time stamp Traditional data mining techniques cannot be applied to temporal data. This necessitates that we reconsider the existing data mining techniques.

The offline data-mining techniques should be able to learn the trends in the overall time intervals and the relationships among features (or attributes) over time intervals for classification under considerations. We also propose to develop temporal rule induction algorithms that learn rules from temporal data.

We propose to develop temporal data mining and fusion that are adaptable to patients' special status, such as emergency state or special drug dose. Using the adaptable temporal mining and fusion, the rules more relevant to the special status of individual patients can be discovered.

3.4 System Evaluation

The system performance will be evaluated by accuracy, efficiency and time delay of diagnosis by comparing the decision making made by medical doctors and new medical query systems. Specificity and sensitivity will be mapped onto the plot of recover operating characteristics (ROC) curves to show overall system performance. The area calculation under the ROC will be used as the performance index of our system.

4. Conclusion

The paper proposed an infrastructure of stream data management, mining and fusion for monitored patients. The infrastructure consists of online and offline engines and temporal databases and data warehousing. The online engine also fuses data from multiple sources. The project for the implementation of the proposed infrastructure is ongoing project. The proposed infrastructure will be implemented in the ICU at Drexel Medical School and able to improve the quality of care for patients in Intensive Care Units (ICUs).

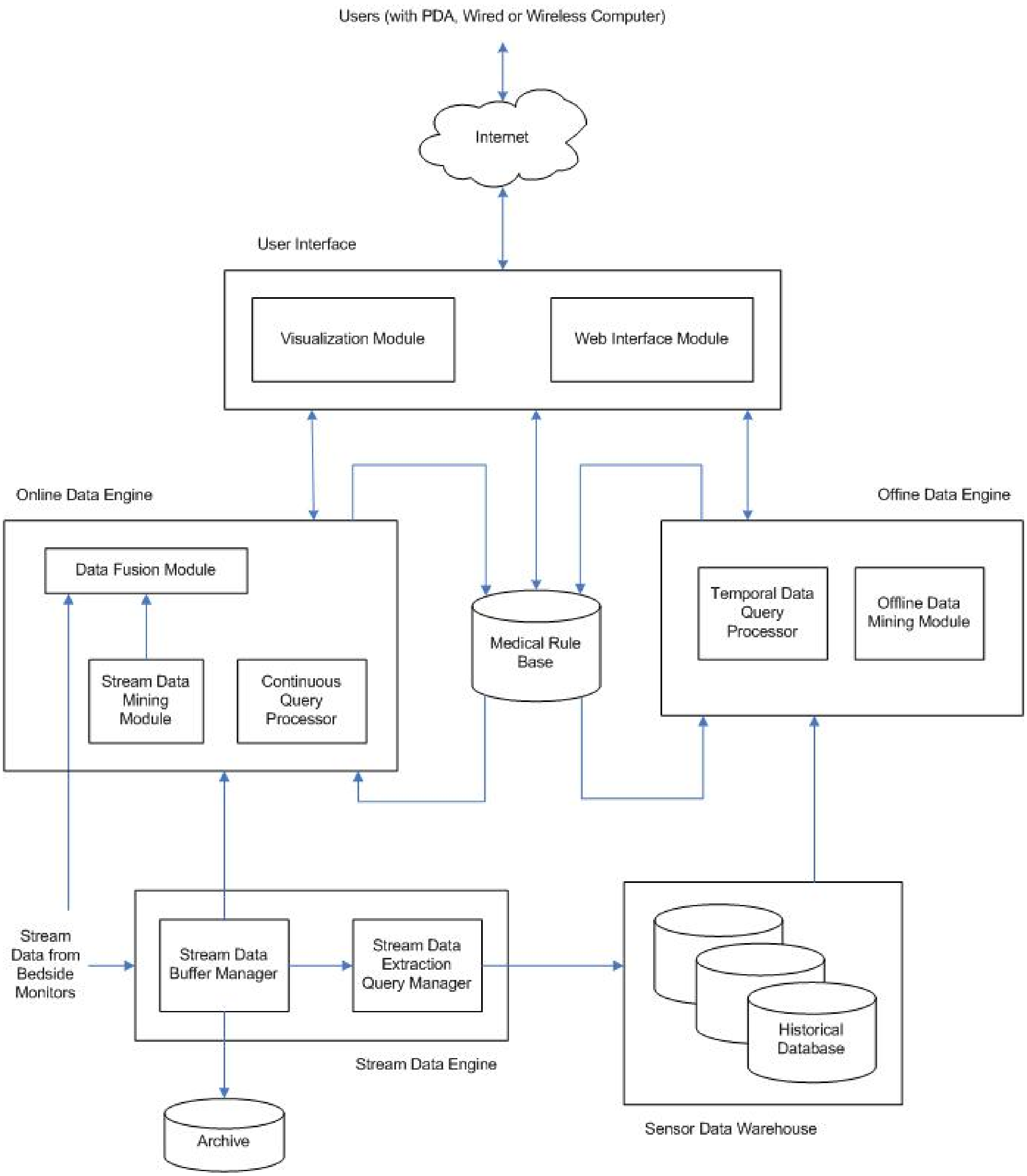

Figure 1. The infrastructure for Stream Data Mining, Fusion, and Management (StreDaMiMa)

REFERENCES

[1] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom, "Models and issues in data stream systems," presented at PODS, 2002.

[2] P. Bonnet, J. Gehrke, and P. Seshadri, "Towards Sensor Database Systems," presented at Int'l Conference on Mobile Data Managment, 2001.

[3] D. Abadi, D. Carney, U. Cetintemel, M. Cherniack, C. Convey, C. Erwin, E. Galvez, M. Hatoun, J. Hwang, A. Maskey, A. Rasin, A. Singer, M. Stonebraker, N. Tatbul, Y. Xing, R.Yan, and S. Zdonik, "Aurora: A Data Stream Management System (Demonstration),," presented at ACM SIGMOD, 2003.

[4] A. Bulut and A. K. Singh, "A Unified Framework for Monitoring Data Streams in Real Time," presented at 21st International Conference on Data Engineering, Tokyo, Japan, 2005.

[5] A. F. Cardenas, R. K. Pon, and R. B. Cameron, "Management of Streaming Body Sensor Data for Medical Information Systems," presented at International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS '03), Las Vegas, Nevada, 2003.

[6] S. Sharshar, L.Allart, and M.-C. Chambrin, "A New Approach to the Abstraction of Monitoring Data in Intensive Care," presented at 10th Conference on Artificial Intelligence in Medicine, AIME 2005, Aberdeen, UK, 2005.

[7] P. Terenziani, R. T. Snodgrass, A. Bottrighi, M. Torchio, and G. Molino, "Extending Temporal Databases to Deal with Telic/Atelic Medical Data," presented at 10th Conference on Artificial Intelligence in Medicine, AIME 2005, Aberdeen, UK, 2005.

[8] P. J. F. Lucas, N. C. d. Bruijn, K. Schurink, and A. Hoepelman, "A Probabilistic and Decision-Theoretic Approach to the management of Infectious Disease at the ICU," *Artificial Intelligence in Medicine*, vol. 19, pp. 251-279, 2000.