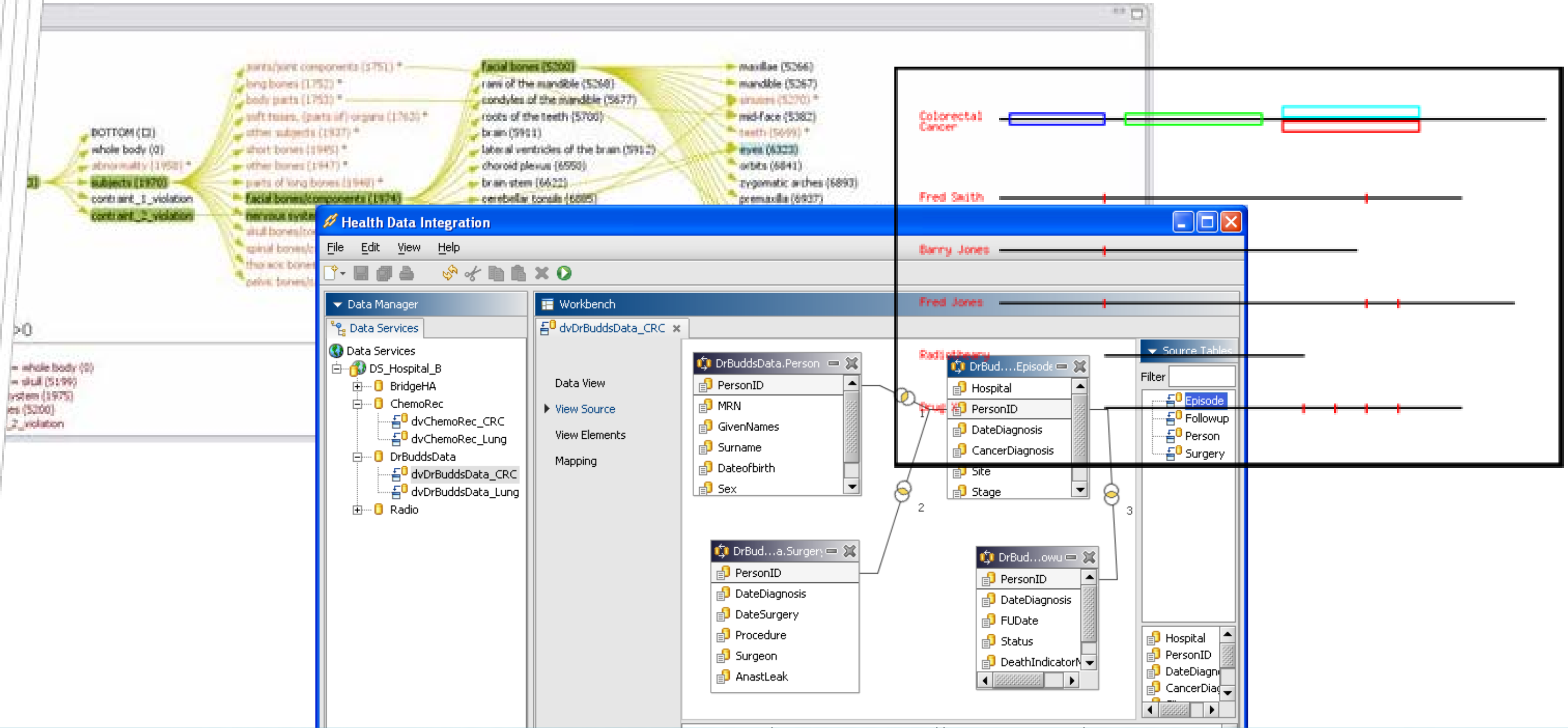


www.csiro.au

Real-Time Analysis of Clinical Time Series Data

Q. Zhang*, C. Pang*, D.P. Hansen*, C. Cheung°, M. Steyn°

* The Australian E-Health Research Centre, CSIRO ICT Centre, Brisbane, Australia

° The Anaesthesia RBWH Qld, Australia

CSIRO

Introduction

- **Physiological time series data has been collected to record details of the physiological state of patients.**
  - High Frequency Signals (ECG, EEG)
  - Low Frequency Signals (Heart Rate, Oxygen Saturation)
- **Analyses of these signals are *vital* for patient safety and management.**
- **Traditional approaches are[1]:**
  - Sample and Store the data in a RDBMS
  - Analyze the data offline

Challenges and Approaches

- **Challenges:**
  - Low sampling rate makes inaccurate records of high frequency data:
    - Uniform sampling is a pervasive method used in ECG data display and analyses.
    - Low sampling rate can seriously distort the original data
    - High sampling rate can cause unnecessary storage overheads since most parts of ECG are almost identical.
- **Our Approaches[3]:**
  - Compress data intelligently using wavelet techniques
  - Guarantee the approximation quality through building error-bound wavelet synopses

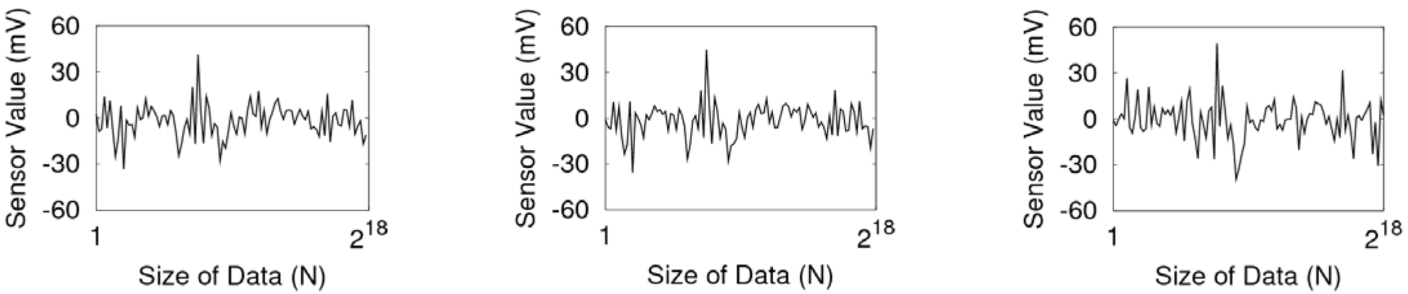
Challenges and Approaches [cont]

- **Challenges:**
  - Large amount of data makes real-time analyses impossible:
    - Complicate queries downgrade whole system performance.
    - Heavy and frequent I/O accesses cause long response time.
- **Our Approaches:**
  - Query directly on the compressed data, i.e. the synopses
  - We can thus save CPU and I/O costs due to the small size of synopses

E-HEALTH RESEARCH CENTRE Queensland Government Preventative Health NATIONAL RESEARCH FLAGSHIPS CSIRO

Techniques – building synopses

- Our compressing techniques:
  - based on the digital wavelet transformation (DWT)
  - but can compress the data under a given error bound
  - and also can build the synopses in one pass processing of the original data
  - time complexity $O(N)$
  - space complexity $O(\lg N)$



(a) Original EEG Data **(b) Approx. EEG by DWT** **(c) Approx. EEG by Sampling**

Compression rate: 0.6%

E-HEALTH RESEARCH CENTRE
Queensland Government
Preventative Health
NATIONAL RESEARCH FLAGSHIPS
CSIRO

Techniques – query wave patterns

- ECG data is highly periodic time series data, termed as pseudo-periodic time series data:
    - can be partitioned into waves that generally have similar durations
    - the shapes of consecutive waves are highly similar
    - a change in the shape is considered significant

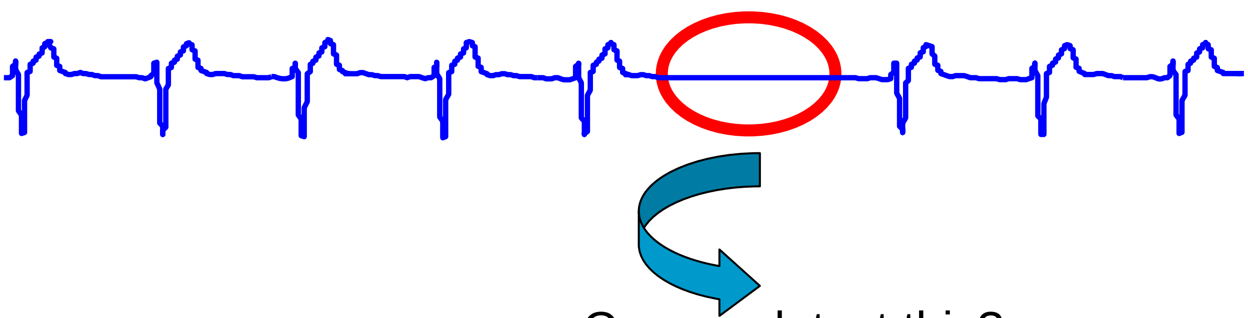
- We thus focus on analysing and mining wave shapes [2]:
    - Classification
    - Indexing
    - Clustering
    - Abnormal Detection
    - Rule Detection

Our Research Objectives:

- **Alerts and anomaly detection in real time (online) from signals:**

Pattern analysis in real time.

Require efficient query algorithms



Can we detect this?

EHEALTH RESEARCH CENTRE

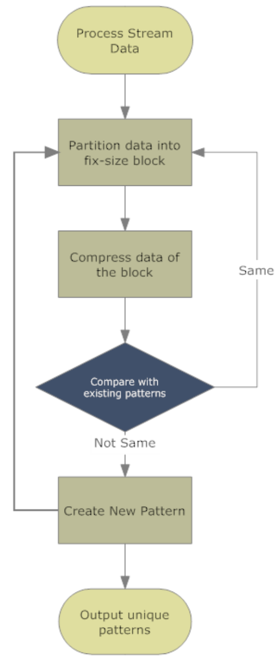
Queensland Government

Preventative Health NATIONAL RESEARCH FLAGSHIPS

CSIRO

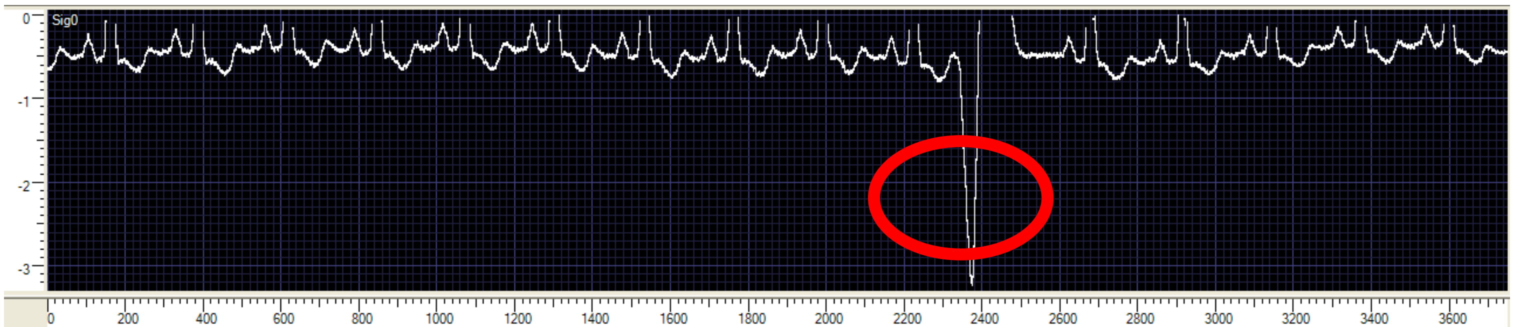
Prototype of Abnormal Pattern Detection

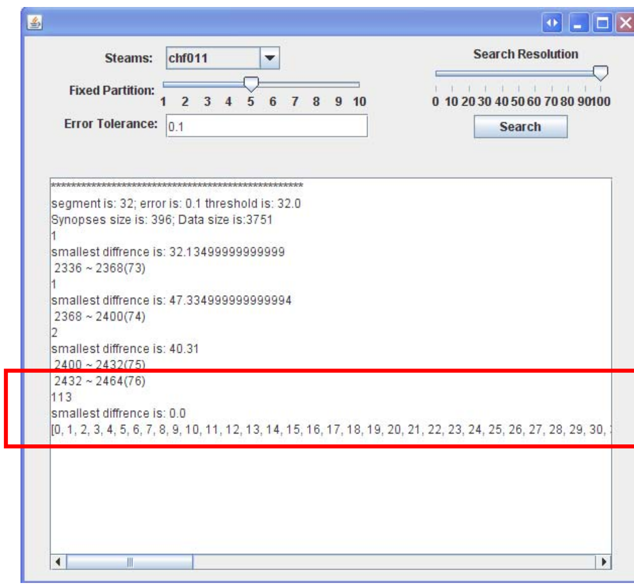
- We build a prototype to detect abnormal patterns:
  1. Partition data into fixed size blocks
  2. Build a synopsis for each block
  3. Compare synopses to judge whether two patterns are similar



E HEALTH RESEARCH CENTRE
Queensland Government
Preventative Health
NATIONAL RESEARCH FLAGSHIPS
CSIRO

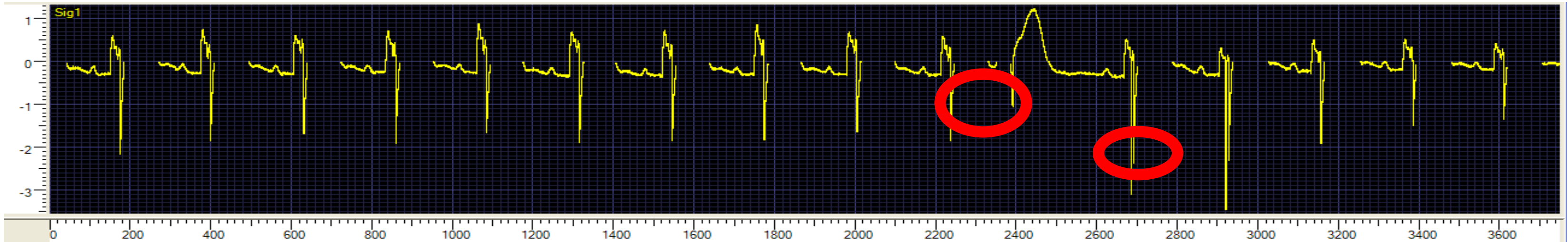
Example 1: pattern detection on ECG

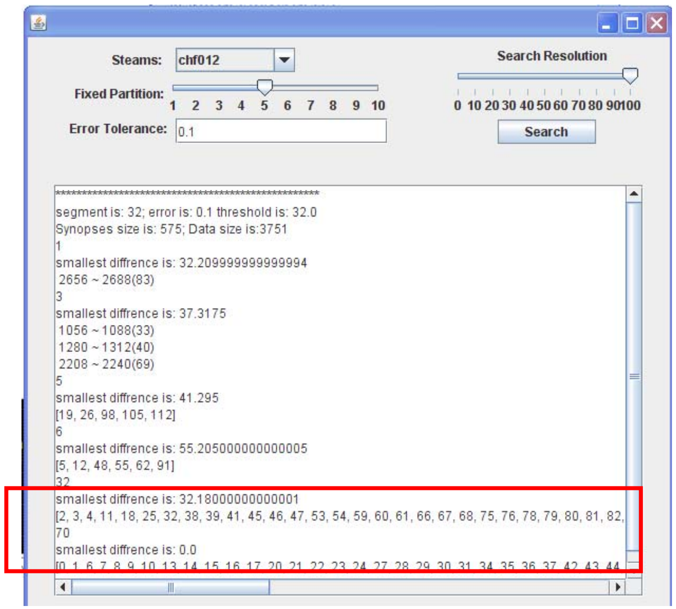




Data size: 3751
Synopses size: 396 (90% compress rate)
Detected abnormal patterns from:
2336 ~ 2464

Example 2: pattern detection on ECG





Data size: 3751
Synopses size: 575
(85% compress rate)
Detected abnormal patterns from:
2656 ~ 2688
2208 ~ 2240

E-HEALTH RESEARCH CENTRE
Queensland Government
Preventative Health
NATIONAL RESEARCH FLAGSHIPS
CSIRO

Conclusions

- **Compressed data size:**
  - our system takes only 1% space of the original data to store the synopses
  - And it takes less than 1 second to build a synopsis

- **Variations on patterns:**
  - many variation can be detected correctly but robustness needs to be enhanced.

- **In the future, we will investigate:**
  - if it is possible to construct only one synopsis with variations to describe those similar patterns.
  - if the variable size block works better than fixed size
  - how to answer other types of queries as listed before

References

[1] Moody GB, Mark RG. 'A database to support development and evaluation of intelligent intensive care monitoring' *computers in Cardiology 1996;23:657-660*

[2] Jessica Lin, Eamonn Keogh, Stefano Lonardi, Bill Chiu. 'A symbolic representation of time series, with implications for streaming algorithms' in proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego. CA. June 13 2003

[3] C. Pang, Q. Zhang, D. Hansen, A. Maeder. 'Constructing Unrestricted Wavelet Synopses under Maximum Error Bound', technical report 08/209, Australia e-Health Research Center