Data mining on temporal data: A visual approach and its clinical application to hemodialysis

Luca Chittaro, Carlo Combi, Giampaolo Trapasso
HCl Lab, University of Udine, Italy

By: Abdulla Alasaadi, Hany M. SalahEldeen
CS 796/896 - Spring 2011, Old Dominion University
About the Paper

- **Authors:** Luca Chittaro, Carlo Combi, Giampaolo Trapasso
- **Affiliation:** HCI Lab, University of Udine, Italy.
- **Venue:** Visual Languages and Computing
- **Year:** 2003.
- **Software:** Interactive Parallel Bar Charts [IPBC].
Introduction

The quantity and complexity of data stored in clinical databases by automated medical devices is rapidly and continuously increasing.

Provides clinicians with easy-to-use interactive tools to analyze huge amounts of this data.

Purposes:
- supporting clinical decision making,
- evaluating the quality of the provided care,
- and carrying out medical research.
Introduction

Some techniques for visual data mining (VDM) of multidimensional clinical databases are illustrated in [2] [3].

Introduction

A different approach is presented by [3] and is based on tables displaying records and their attributes in **highly compressed format** such that they fit onto the screen.

Compressed table mode.
• Based on bar charts, focusing on clinical databases containing time-series data.

• Familiarity is an important requirement.

A clinical VDM system has to achieve two possibly conflicting goals:
(i) Offering powerful data analysis capabilities, while
(ii) Minimizing the number of concepts and functions to be learned by clinicians.
The hemodialysis domain

A medium-sized center collects more than 228 millions of parameter values per year

- It is very important for clinicians to be able to evaluate the quality of:
  (i) each single hemodialysis session,
  (ii) all the sessions concerning the same patient, and
  (iii) sets of patient sessions concerning a specific hemodialyzer device or a specific day.

- Early detection of problems in the quality of the hemodialytic treatment.
The Proposed Approach

The System: Interactive Parallel Bar Charts (IPBC)

- Connects to the hemodialysis clinical database,
- produces a visualization that replaces tens of separate screens used in traditional hemodialysis systems, and
- extends them with a set of interactive tools.
- The visualizations are based on bar charts.
Fig. 1. A Parallel Bar chart with a pop-up round toolbar.
• 3D space can significantly increase both the number of time-series and the number of values associated.

Problems:
• Occlusions,
• 3D navigation,
• Difficulties in comparing heights,
• Proper use of space,
• The need for effective interaction techniques to aid the user in the analysis of large data sets (e.g. highlighting interesting patterns, checking trends, y).
The RoundToolbar (RT)

- Solved The problems of:
  - The possible waste of screen space towards the corners of the screen.
  - Traditional menus: require lengthy mouse movements

- Upper-half: how bar charts are viewed.
- Lower-half: provide different kinds of visual exploration tools.

Fig. 2. Controls in the View mode
Changing viewpoint

Fig. 3. Viewpoint movements: (a) high–low, (b) rotate, (c) near–far.
Temporally grouping data

• Visualize the mean of parameter values grouped over intervals of a given duration.

• Group by values - allows the user to visually identify groups of consecutive samples with the same value.
Fig. 4. Mean function applied to data grouped over 30 min intervals.
Fig. 5. Grouping data by value.
Managing occlusions

• Using the 2D/3D mode on the RT, the user can transform the parallel bar chart into a matrix format and vice versa.

• **Collapse series**: this typically results in a flattening of portions of the series.

• **Collapse**: this typically results in a flattening a specific time-series.
Fig. 6. Matrix visualization.
Fig. 7. Removing occlusions by flattening entire time-series.
Fig. 8. Removing occlusions by flattening portions of time-series.
Hiding options

The Hide mode allows the user to display/hide some information for parallel bar charts.

The user can display/hide:
(i) The grid on the plane identified by the time axis and the axis for time-series.
(ii) The result of each hemodialysis.

Successful sessions are visually identified by a filled green circle displayed at the beginning of the corresponding bar chart, while unsuccessful sessions are identified by a filled red circle.

Clinical events: blue signs on top of bar charts in the correct position [staff intervention during a session].
Fig. 9. Visualizing/hiding details (Hide mode).
Fig. 10. Adding space between series.
Dynamic queries
Comparing data with (time-varying) thresholds

• How many and which values are below or above a given threshold.

• Dynamic queries can be used when the threshold is constant.

• The required threshold is often time-varying.

• **The Tide mode** adds a semitransparent solid to the visualization [Linear Tide].
Comparing data with (time-varying) thresholds

Fig. 12. Tide mode.
Pattern matching

A SQL-like query language named Simple Query Language for Time-series (SQL-TS) is proposed with specific constructs for defining sequential patterns to match against the time-series database

```sql
SELECT S1.SessionID, S1.SampTime
FROM AllSeries
CLUSTER BY (PatientID, SessionID)
SEQUENCE BY SampTime
AS (S1, S2, y, Sm, Sn)
WHERE
S1.PatientID = ConsideredPat AND
ABS(S1.ParName  U1) p ABS(S1.ParName)*TV/100 AND
....
```
http://hcilab.uniud.it/demos-videos/item4/vedi.avi

http://cs.odu.edu/~aalasaad/CS896/vedi.avi
What is Data Mining?

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information.
Mining Hemodialyctic data

- Gathering real clinical task data that clinicians gather.
- This analysis leads to increase the quality of care to each patient.
- Several parameters are obtained, each on a time-series. Among which are:
  - Systolic blood pressure.
  - Diastolic blood pressure.
  - Heart rate.
  - Blood Volume (Flow in hemodialyzer QB)
  - ...etc.
Mining Hemodialyctic data

Dialysis sessions

For each parameter
ex: Systolic blood pressure

Time frames
Mining Hemodialyctic data

Data mining could be used in:
1. Mining patient signs data.
2. Mining blood volume data.
3. Mining related clinical parameters.
4. Mining for similar patterns.
1- Mining patient signs data

- Remove physiologically out of range data.
- Examine color branded data to find undesired situations.
- Try to find therapeutic actions over time that had this effect to fix next sessions.
2- Mining blood volume data

- Percentage of blood volume reduction due to remove water in dialysis.
- Patient’s blood pressure decreases.
- Use the Tide mode.
2- Mining blood volume data

- Can use grouped Tide mode to study trends

5 min interval grouping.
3- Mining related clinical params.

- Test QB with blood pressure.
- QB reduced by nurses when pressure is suboptimal.
- Used to check if suboptimal QB is due to low blood pressure.
4- Mining for similar patterns

- Select the region of suboptimal results (QB).
- Tune the tolerance parameter.
- Extract patterns.
4- Mining for similar patterns
Evaluation of IPBC

- With clinical staff of hemodialysis center in Hospital Mede in Pavia Italy.
- Iterative prototyping to determine:
  • not easy to understand features.
  • new required functionalities.
- Development went through several phases.
Phases of Development

1- VRML based visualizer based on time series collections.
   - peaks and valleys.
   - accepted visualization in science.
   - easy to distribute and visualize files on internet. (plugins, web-browser addons).
     - FAILED to meet needs of end users because:
       • Difficult to use by end users.
       • Lead to disorientation and occlusion.
       • Not familiar to clinicians.
       • Lacking functionalities.
Phases of Development

1- VRML based visualizer based on time series collections.
Phases of Development

2- First prototype to IPBC.
   - Developed by OpenGL for more power and manipulation.
   - Advantages:
     - quickly learnt and remembered.
   - Disadvantages:
     - need usability improvements.
Phases of Development

3- Final IPBC.
   - Developed from all feedback of clinicians
   - Disadvantages:
     - screen usage with increased number of related collections.
     - most times when inspecting a time-series don’t need all time-series of other parameters, need just few values.
Integrating parallel bar charts with parallel coordinate plots

- Due to last disadvantage.
- Integration of IPBC with well-known Parallel Coordinate Plots (PCP)
- While inspecting the bar chart other parameters appear on PCP on same screen.
- Each parameter is a separate vertical axis.
- Line connecting values on axes.
Integrating parallel bar charts with parallel coordinate plots

Each bar corresponds to a line
Integrating parallel bar charts with parallel coordinate plots

- Could be used also in pattern matching
Integrating parallel bar charts with parallel coordinate plots

- Could be used in ANIMATE mode
- Select sequence of bars
- Show animation of lines in PCP corresponding to evolution of parameters.
- Play, Stop, Rewind, Forward, Variable Speed Capabilities.
- Benefits: Detect anomalies in evolving parameters.
- this task is familiar to clinicians because it is done repeatedly in echographies and angiographies.
Integrating parallel bar charts with parallel coordinate plots

- Could be used in ANIMATE mode
Future Work

- Possible integration with other visualizations and calculations (medical literature quality index per session).
- Extend IPBC to deal with different abstraction levels. (to detect trends and temporal patterns AND fine exploration of time series).
- More advanced pattern matching techniques for extended queries.
Proposed Questions

1. Why didn’t they propose overlaying several IPBC on top of each other.
2. Why didn’t they apply the non-linear Tide?
3. Why there is no surfacing capabilities? (sometimes I need approximation not real data).
4. Why no prediction capabilities?
Questions?