Interactive Visual Analysis of Complex Data: Introduction, Families of Curves

Krešimir Matković
Vienna University of Technology,
VRVis Research Center, Vienna
Overview 1

Introduction – summary once more

Coordinated Multiple Views – summary once more

Complex data: families of curves
- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data
Multidimensional Multivariate Data

Conventional approach deals with n-dim. Euclidian spaces
Each item is a point in n-dim. Space (n-tuple)

\[ Pi = [x_1, \ldots, x_i, \ldots, x_n] \]

<table>
<thead>
<tr>
<th>STATION</th>
<th>AVERAGE TEMP</th>
<th>PRESSURE</th>
<th>ELEVATION</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>1015</td>
<td>200</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1020</td>
<td>300</td>
<td>...</td>
</tr>
</tbody>
</table>
Coordinated multiple views

Show various dimensions simultaneously

Use mostly simple 2D views

Interaction:
- Linking & brushing
- Select items in one view and they will be highlighted in all views
Complex Data

Engineering data does not always fit to this data model

Typical simulation:
- N control parameters
- n simulation runs \((n \leq 2^N)\)
- M output values which can be mappings \((f(x), x\) is time, crankangle,\(\ldots\))

Point \(P_i\) is the same, but \(x_i\) can be a mapping
Mapping as a dimension 1

Such data model is common

Simulations, Meteorology, Traffic, Telecommunication, Financial data, ...

Such mappings themselves are common

Often described as 2D Euclidian space

Our approach differs from conventional

- Mapping is a dimension of the space
Mapping as a dimension 2

<table>
<thead>
<tr>
<th>STATION</th>
<th>AVERAGE TEMP</th>
<th>PRESSURE</th>
<th>ELEVATION</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>1015</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1020</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STATION</th>
<th>TEMPERATURE over the year</th>
<th>PRESSURE</th>
<th>ELEVATION</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Jan,5);(Feb,8);</td>
<td>1015</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(Jan,6);(Feb,10);</td>
<td>1020</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>
Curve View

If there are $N$ points $P_i$

And dimension $k$ is a mapping

Each $x_k$ is a function graph

There are $N$ function graphs for dimension $k$
  – A Family of Function Graphs

Curve View displays such a family

Interaction

Curve view in a multiple linked view setup

Need for efficient brush

Line brush

- Selects all curves which cross the line
- Very intuitive
- Easy to define
- Would be much more complex using SQL
Overview 1

Introduction – summary once more

Coordinated Multiple Views – summary once more

Complex data: families of curves
- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data
Typical Visualization Tasks

- Exploration
  - find the unknown, unexpected
  - hypothesis generation
- Analysis
  - confirm or reject hypotheses
  - information drill-down
- Presentation
  - communicate/disseminate results

Traffic Sensor Network (Minneapolis, St. Pauls)

- 12 weeks (84 days)
- 564 sensors
- daily data, aggregated from measurements all 30 secs.
- \(2 \times 47376 = 94752\) graphs
  - (144 \(f(t)\)-values each, one per 10 mins.)

[TVCG 2006]
Typical Visualization Tasks

- Exploration
- Hypothesis generation
- Analysis
- Confirm or reject hypotheses
- Information drill-down
- Presentation
- Communicate/disseminate results

... 2 sensors, 2 days (Sun+Mon)!

[TVCG 2006]
Typical Visualization Tasks

- Exploration
- Hypothesis generation
- Analysis
- Confirm or reject hypotheses
- Information drill-down
- Presentation
- Communicate/disseminate results

... 293 outliers (out of 47376 ≈ 0.6%)!

... negative(!) volume-values brushed...

[TVCG 2006]
Typical Visualization Tasks

- Exploration
- Hypothesis generation
- Analysis
- Confirm or reject hypotheses
- Information drill-down
- Presentation
- Communicate/disseminate results

[TVCG 2006]
Typical Visualization Tasks

- Exploration
- Find the unknown, unexpected
- Hypothesis generation
- Analysis
- Confirm or reject hypotheses
- Information drill-down
- Presentation
- Just 1 outlier!

... again the same sensor!

... just 1 outlier!

[TVCG 2006]
Typical Visualization Tasks

- Exploration
  - find the unknown, unexpected
- Hypothesis generation
- Analysis
  - confirm or reject hypotheses
- Information drill-down
- Presentation
  - communicate/disseminate results

whole day no cars?!

why (just) here???

[TVCG 2006]
Overview 1

Introduction – summary once more

Coordinated Multiple Views – summary once more

Complex data: families of curves

- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data

Motivation

Simulation

- widely used in science and engineering
- helps in understanding various phenomena
- recent trend:
  - multiple simulation runs
  - variation of various control parameters
  - study influence on output
  - understanding of simulation model
  - understanding of physical phenomena

Simulation and Visualization 1

Common practice:
- define a model
- run simulation
- visualize results
- change the model
- run simulation ...

Krešimir Matković
InfoVis, TU Wien, 25. 04. 2012
Simulation and Visualization 2

Multiple runs:
- define a model
- define control parameters
- run simulation
  - for each combination of control parameters
  - for some of the combinations (steering)
- use interactive visual analysis
  - to understand (optimize) the model
  - to understand the phenomena

Krešimir Matković
InfoVis, TU Wien, 25. 04. 2012
Interactive Visual Steering

Model, parameters, many simulation runs

How many is many?
- 11 parameters / 10 variations each
- $10^{11}$ combinations
- 10 simulations per minute (very fast) - $10^{10}$ minutes
- 19 000 years!

Basic Idea

Start with a simple model. Assume the rest to be ideal. Expand the model. Tune the expanded model. Expand the model… It is always possible to go back!
Interactive Steering

3 loops
- Explore initial results set
- Change parameter values
- Change model

Tabular data
- Initial table
- Add / remove rows
- Add / remove columns

Interaction

3 loops

- A – real time interaction
- B – “live” parameter refinement
  (A loop functioning)
- C – model change - after initial runs and data management update
  same as B
First Step

Start with a simple model
Assume the rest to be ideal
Tune the first model
4 parameters, 750 combinations
12 minutes
Two targets

Krešimir Matković  
InfoVis, TU Wien, 25. 04. 2012
Interactive Visual Analysis

Second Step

Second Step

Different responses

H 210
N 1.9
S 2.9

CV_size 10
Z_inl 1.0
Z_out 2.0

Parameter refinement

Additional investigation needed

– refine parameters
Output at various steps

Krešimir Matković
InfoVis, TU Wien, 25. 04. 2012
Final model

Everything is set now
We wanted to check all output curves
Actuator variations
Interesting peaks detected!

Final solution

One operation point
Choose one and test
If everything OK – done!

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Name</th>
<th>Final value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>d_sac</td>
<td>0.75 mm</td>
</tr>
<tr>
<td>I_2</td>
<td>Alpha_seat</td>
<td>50 degrees</td>
</tr>
<tr>
<td>I_3</td>
<td>C_turb</td>
<td>0.9</td>
</tr>
<tr>
<td>I_4</td>
<td>mju</td>
<td>0.6</td>
</tr>
<tr>
<td>II_1</td>
<td>CV_size</td>
<td>10 mm³</td>
</tr>
<tr>
<td>II_2</td>
<td>Z_inl</td>
<td>1.6</td>
</tr>
<tr>
<td>II_3</td>
<td>Z_out</td>
<td>2.6</td>
</tr>
<tr>
<td>III_1</td>
<td>Res_bypass</td>
<td>2.0</td>
</tr>
<tr>
<td>III_2</td>
<td>Res_Outlet</td>
<td>1.0</td>
</tr>
<tr>
<td>III_3</td>
<td>Area_Bypass</td>
<td>0.032 mm²</td>
</tr>
<tr>
<td>III_4</td>
<td>Area_Valve</td>
<td>0.07 mm²</td>
</tr>
<tr>
<td>IV_1</td>
<td>HPP_Length</td>
<td>300 mm</td>
</tr>
<tr>
<td>IV_2</td>
<td>RV_size</td>
<td>30 cm³</td>
</tr>
</tbody>
</table>
Overview 1

Introduction – summary once more

Coordinated Multiple Views – summary once more

Complex data: families of curves
  - Traffic Surveillance
  - Engineering
    - Interactive visual steering
    - A hybrid approach to visual steering
  - Bio-Sygnals and ICU Data
    - Heart reinervation
    - IVA of ICU Data
Complex system

A complex system

4 injectors

AVL BOOST HydSim

Many parameters
Main Goals

To understand the model

To understand the phenomena

To optimize the model for a given operational regime

- depending on load and goal
- the shape of injection curve is essential criterion
- pilot and main injection

Possible approaches

Automatic – use automatic optimization methods to determine optimum point of operation

- a regression model is needed

Our approach: Combine automatic and interactive approach

- human in the loop needed
- time consuming
- besides optimum user gets insight in the system operation
Interactive Visual Steering

Computational steering and interactive visualization

- simulation and visualization often decoupled
- integrated system is much more powerful
- integration of: modeling, simulation, visual analysis
- run simulation
- explore results
- change the simulation model or parameters from visualization
Multiple Simulation Runs

Needed for interactive and automatic approach
Define a model
Define control parameters
Run simulation
  - for each combination of control parameters
  - for some of the combinations
Use interactive visual analysis
  - to understand (optimize) the model
  - to understand the phenomena
Data 1

Scalar parameters define each simulation model block

Blocks are connected in a model

Simulation computes state variables for each block

State variables are mostly time (crank-angle) dependant

We do not have a conventional data with a multidimensional point

| Run | Parameter 1 | Parameter 2 | ... | Output1(t) | Output2(t) | ...
|-----|-------------|-------------|-----|------------|------------|-----
| 1   | 10          | 20          |     |            |            |     |
| 2   | 10          | 30          |     |            |            |     |
| 3   | 10          | 40          |     |            |            |     |
| ... | ...         | ...         |     |            |            |     |
Data 2

Optimization methods and regression model building methods expect a multidimensional point

We need to parameterize curves, to represent them with a set of scalars

Impossible to capture all possible variations

Many parameters – good approximation, complex optimization

Interactive methods can use aggregates as well
Basic Idea

Define a model

Define control parameters

Determine design points – combinations of control parameters

Run simulations for initial set of points (a sparse set in the continuous parameter space)

Compute a regression model – needed for automatic optimization

Use interactive visual analysis to explore results:
  - identify a region of interest
  - specify constraints
  - automatically find an optimum in a selected region using the regression model
  - run simulation for the optimum and for points in neighborhood

Illustrative example – initial data

Six parameters varied – 2700 simulation runs

Regression model built
Interactive visual analysis used to explore the data

First goals

We want balanced system and balanced individual injectors

Some of the goals:

- small difference between injection pressures of individual pilot injections for each cylinder
- small difference between amounts of fuel injected of individual pilot injections for each cylinder
- as fast as possible needle opening and closing velocities for pilot and main injections
- etc.
Brushing the goals
Setting Constraints

Based on the selection user sets constraints

“Orders” an optimum from automatic system

Optimum is based on the regression model, there are no curves
The optimum is computed

Load the optimum

Run simulation for optimum points – get full data

Compute points in the neighborhood – optimum might be wrong, it's based on the regression

Explore the neighborhood

Explore the neighborhood

P3_pt_diff was minimized

There are better points in the neighborhood!
Another example – changing the regression model

Final decision – a compromise
Conclusion

Close coupling of simulation, optimization, and visualization

Complex systems – complex data

- too complex for pure automatic or pure interactive approach

An integrated system and workflow presented

Supports experts in understanding the system, automatic optimization speeds up the process

Very positive feed-back from engineers

Overview 1

Introduction – summary once more

Coordinated Multiple Views – summary once more

Complex data: families of curves

- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data

Heart Reinervation 1
Heart Reinervation 2

Overview 1

Introduction – summary once more

Coordinated Multiple Views – summary once more

Complex data: families of curves

- Traffic Surveillance
- Engineering
  - Interactive visual steering
  - A hybrid approach to visual steering
- Bio-Sygnals and ICU Data
  - Heart reinervation
  - IVA of ICU Data
ICU Data

Modern Intensive Care Unit

- a lot of high technology equipment
- routinely logging of
  - many physiological measurements
  - laboratory results
  - interventions information
- these information can help intensive care physicians to

understand physiological, pathological and therapeutic processes
and so, consequently, to improve medical care
Usage of Data

Huge amounts of data available

Great potential but:
- many proprietary formats, data bases rarely merged
- statistical models are complex and hard to develop without merged DBs
- communication between medical and statistics experts

State of the art
- various regression models and advanced statistics
- Protocols (researchers – daily practice)
- APACHII Score, for example

Interactive Visual Analysis
- novel concept for ICU medical researchers

Krešimir Matković
InfoVis, TU Wien, 25. 04. 2012
Data 1

~1500 patients

<table>
<thead>
<tr>
<th>Time</th>
<th>Patient</th>
<th>Age</th>
<th>apa</th>
<th>gsc</th>
<th>temp</th>
<th>inr</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:00</td>
<td>4965</td>
<td>77</td>
<td>23</td>
<td>7</td>
<td>37,6</td>
<td>1,3</td>
</tr>
<tr>
<td>01:00</td>
<td>1404</td>
<td>43</td>
<td>18</td>
<td>12</td>
<td>40,4</td>
<td>1</td>
</tr>
<tr>
<td>01:00</td>
<td>2344</td>
<td>73</td>
<td>24</td>
<td>12</td>
<td>38,2</td>
<td>1,2</td>
</tr>
<tr>
<td>02:00</td>
<td>4965</td>
<td>77</td>
<td>23</td>
<td>7</td>
<td>37,4</td>
<td>1,3</td>
</tr>
<tr>
<td>02:00</td>
<td>1404</td>
<td>43</td>
<td>18</td>
<td>12</td>
<td>40,5</td>
<td>1</td>
</tr>
<tr>
<td>02:00</td>
<td>2344</td>
<td>73</td>
<td>24</td>
<td>12</td>
<td>38,5</td>
<td>1,2</td>
</tr>
<tr>
<td>04:00</td>
<td>4965</td>
<td>77</td>
<td>23</td>
<td>7</td>
<td>37,5</td>
<td>1,3</td>
</tr>
<tr>
<td>04:00</td>
<td>1404</td>
<td>43</td>
<td>18</td>
<td>12</td>
<td>40,4</td>
<td>1</td>
</tr>
<tr>
<td>04:00</td>
<td>2344</td>
<td>73</td>
<td>24</td>
<td>12</td>
<td>38,6</td>
<td>1,2</td>
</tr>
</tbody>
</table>

Krešimir Matković

InfoVis, TU Wien, 25. 04. 2012

complex processing and analysis

complex processing and analysis
Coordinated Multiple Views (CMV)

Krešimir Matković
InfoVis, TU Wien, 25.04.2012
nD space

Conventional approach deals with n-dim. Euclidian spaces

Each item is a point in n-dim. Space (n-tuple)

\[ P_i = [x_1, \ldots, x_i, \ldots, x_n] \]
Complex Data

Medical data does not always fit to this data model

- We want data that belongs to one patient to be in one record
  - more natural way of thinking of data
  - new possibilities of analysis

Point $P_i$ is the same, but $x_i$ can be a mapping

Mapping can generally be $f(x), f(x,y), set, ...$

$P_i = [x_1, \ldots, x_i, \ldots, x_n]$

[Konyha et al. ’06, Matkovic et al. 08, 09, 10, Freiler et al. 08, ...]
Family of Curves

<table>
<thead>
<tr>
<th>Time</th>
<th>Patient</th>
<th>Age</th>
<th>apa</th>
<th>gsc</th>
<th>temp</th>
<th>inr</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:00</td>
<td>4965</td>
<td>77</td>
<td>23</td>
<td>7</td>
<td>37,6</td>
<td>1,3</td>
</tr>
<tr>
<td>01:00</td>
<td>1404</td>
<td>43</td>
<td>18</td>
<td>12</td>
<td>40,4</td>
<td>1</td>
</tr>
<tr>
<td>01:00</td>
<td>2344</td>
<td>73</td>
<td>24</td>
<td>12</td>
<td>38,2</td>
<td>1,2</td>
</tr>
<tr>
<td>02:00</td>
<td>4965</td>
<td>77</td>
<td>23</td>
<td>7</td>
<td>37,4</td>
<td>1,3</td>
</tr>
<tr>
<td>02:00</td>
<td>1404</td>
<td>43</td>
<td>18</td>
<td>12</td>
<td>40,5</td>
<td>1</td>
</tr>
<tr>
<td>02:00</td>
<td>2344</td>
<td>73</td>
<td>24</td>
<td>12</td>
<td>38,5</td>
<td>1,2</td>
</tr>
<tr>
<td>04:00</td>
<td>4965</td>
<td>77</td>
<td>23</td>
<td>7</td>
<td>37,5</td>
<td>1,3</td>
</tr>
<tr>
<td>04:00</td>
<td>1404</td>
<td>43</td>
<td>18</td>
<td>12</td>
<td>40,4</td>
<td>1</td>
</tr>
<tr>
<td>04:00</td>
<td>2344</td>
<td>73</td>
<td>24</td>
<td>12</td>
<td>38,6</td>
<td>1,2</td>
</tr>
</tbody>
</table>

The table above shows data for multiple patients over time, with columns for Patient ID, Age, and other measurements. The diagram below illustrates a family of curves for the data.
Interaction

Curve view in a multiple linked view setup

Need for efficient brush

Line brush
- selects all curves which cross the line
- very intuitive
- easy to define
- would be much more complex using SQL

Composite brushing
- iteratively combine brushes using Boolean operations
Workflow

- Prepare the data according to the new data model
- Derive data if necessary
- Domain Experts have to familiarize with the tool
- Add specific features needed by domain experts depending on their tasks
- More complex data exploration and analysis
Data preparation

Prepare the data according to the new data model

Improvements of curve view were necessary:
- Support for different lengths of curves
- Support for gaps
- Small and large gaps
- Depict start-end point
- Fill the missing part (optionally)
- Mouse-over-highlighting

Krešimir Matković
InfoVis, TU Wien, 25.04.2012
Additional Features

Introduction of box-plot into CMV

- experts were used in an interactive setup very powerful overview method
- support for multiple brushes – show brushed records and non-brushed records, and all

Inverse brush
- usual approach – show brushed records, and all

Brush summary
- another summary view

Case Study 1 – Confirming a Hypothesis

We are exploring sodium concentration

- **Hyponatraemia** – sodium too low, \(< 130 \text{ mmol/L}\)
- **Hypernatremia** – sodium too high, \(> 150 \text{ mmol/L}\)

Sodium concentration available as a curve per patient

**Known hypothesis** – patients with hyp or hypernatremia have a higher mortality rate

**Brushing hypernatremia patients**

- Use curve view and a long horizontal brush
- Compute maximum aggregate – curve -> scalar and use conventional views

Case Study 2 – Exploring Variability

We suspect that not only absolute values, but rate of change has influence on mortality.

First derivation aggregate curve – curve > curve

Using standard techniques to explore the change

Hypernatremia with rapid fall very dangerous

Special care has to be taken when reducing sodium level

We also experienced patients with rapid fall or rise but no hyper or hyponatremia

 unexpected finding!
 high sodium level variability increases mortality!

Krešimir Matković
InfoVis, TU Wien, 25. 04. 2012
Case Study 3 – Interactive drill down

New data model makes advanced drill down possible

The timing of fall/rise can be easily specified
Conclusion

Mappings as dimension - very powerful concept

Can be applied almost everywhere

Very positive feedback on line brush from Vis community

Very positive feedback from domain experts in various domains
Thank you!

Special thanks for used materials to H. Hauser, and colleagues from VRVis!