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-Signals 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)

\[ P_i = [x_{i1}, ..., x_{in}] \]

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, ...

Point \( P_i \) is the same, but \( x_i \) can be a mapping

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

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>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>116</td>
<td>1015</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>220</td>
<td>1020</td>
<td>300</td>
</tr>
</tbody>
</table>

Curve View

If there are N points Pi
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-Signals and ICU Data
- Heart reinervation
- IVA of ICU Data

Traffic Sensor Network
(Minneapolis, St. Pauls)
- 12 weeks (84 days)
- 564 sensors
- daily data, aggregated from measurements all 30 secs.
- 2·47376 = 94752 graphs
(144 f(t)-values each, one per 10 mins.)
Typical Visualization Tasks

- Exploration
- Find the unknown, unexpected
- Hypothesis generation
- Analysis
- Confirm or reject hypotheses
- Information drill-down
- Presentation
- Communicate/disseminate results

[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-Signals 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 ...

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

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
Tune the first model
Expand the model
Tune expanded model
Expand the model ...
It is always possible to go back!

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

Interactive Steering

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

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

Interactive Visual Analysis

Second Step

Different responses
High CV size – unwanted behavior
Narrow curves – low injected mass
No pilot injection
Same actuator!

Parameter refinement

Additional investigation needed
– refine parameters

Output at various steps
Final model

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

Final solution

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

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-Signals 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
To optimize the regime
  - depending on load and goal in injection can be significantly different
  - the shape of injection curve is an 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) dependent
We do not have a conventional data with a multidimensional point

<table>
<thead>
<tr>
<th>Run</th>
<th>Parameter 1</th>
<th>Parameter 2</th>
<th>Output 1(t)</th>
<th>Output 2(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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
Determine
Run simulation parameter
Compute a
Use interaction
- Identify
- Specify
- Automate
- Run sir

**Illustrative example – initial data**

Six parameters varied = 3700 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**

Krešimir Matković
InfoVis, TU Wien, 25. 04. 2012
Final decision – a compromise

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 feedback 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-Signals and ICU Data
  - Heart reinervation
  - IVA of ICU Data

Heart Reinervation 1

Introduction – summary once more
Coordinated Multiple Views – summary once more

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

Data 1

~1500 patients

<table>
<thead>
<tr>
<th>Item</th>
<th>Patient</th>
<th>Age</th>
<th>pRd</th>
<th>pAp</th>
<th>spAp</th>
<th>tArm</th>
<th>fPed</th>
<th>rArm</th>
</tr>
</thead>
<tbody>
<tr>
<td>01:00</td>
<td>01:00</td>
<td>77</td>
<td>22</td>
<td>77</td>
<td>37.6</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>02:00</td>
<td>02:00</td>
<td>72</td>
<td>22</td>
<td>22</td>
<td>22.2</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>03:00</td>
<td>03:00</td>
<td>73</td>
<td>22</td>
<td>77</td>
<td>37.6</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04:00</td>
<td>04:00</td>
<td>73</td>
<td>22</td>
<td>77</td>
<td>37.6</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>05:00</td>
<td>05:00</td>
<td>73</td>
<td>22</td>
<td>77</td>
<td>37.6</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- complex processing and analysis

nD space

Conventional approach deals with n-dim. Euclidian spaces
Each item is a point in n-dim. Space (n-tuple)
\[ P_i = [x_1, \ldots, x_n] \]
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 $f(x), f(x,y), \text{set}, ...$

[Konyha et al. '06, Matkovic et al. 08, 09, 10, Freiler et al. 08, ...]

Family of Curves

Interconnection of curves

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

Additional Features
Case Study 1 – Confirming a Hypothesis

We are exploring sodium concentration

Hyponatraemia – sodium too low, < 130 mmol/L
Hypernatremia – sodium too high, > 150 mmol/L

Sodium concentration available as a curve per patient


Case Study 2 – Exploring Variability

Hypernatremia with rapid fall very dangerous
Special care has to be taken when reducing sodium level
We also explored patients with rapid fall or rise but no hyper or hyponatremia


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!