Ch 13/14/15: Reduce, Embed, Case Studies

Paper: TopoFisheye Example Present: Biomechanical Motion

Tamara Munzner
Department of Computer Science
University of British Columbia
CPSC 547 Information Visualization
Week 8 & 10 Oct 2017

http://www.cs.ubc.ca/~tmm/courses/547-17F
Ch 13/14/15: Reduce, Embed, Case Studies
Paper: TopoFisheye
Example Present: Biomechanical Motion

Idiom: cross filtering
• item filtering
• co-ordinated views/controls combined
• all scented histogram bisliders update when any ranges change

System: Crossfilter

http://square.github.io/crossfilter/

http://www.cs.ubc.ca/~tmm/courses/547-17F
Ch 13/14/15: Reduce, Embed, Case Studies
Paper: TopoFisheye
Example Present: Biomechanical Motion

Idiom: histogram
• static item aggregation
• task: find distribution
• derived data
  – new table: bins, values are counts
• bin size crucial
  – pattern can change dramatically depending on discretization
• opportunity for interaction: control bin size on the fly

Idiom: scented widgets
• augmented widgets show information scent
  – cues to show whether value is dwelling down further vs looking elsewhere
• concise use of space: histogram on slider

Idiom: scented widgets
• augmented widgets show information scent
  – cues to show whether value is dwelling down further vs looking elsewhere
• concise use of space: histogram on slider

Idiom: Hierarchical parallel coordinates
• dynamic item aggregation
• derived data: hierarchical clustering
  • encoding:
    – cluster bar with variable transparency: line at mean, width by min/max values
  • color by proximity in hierarchy

Idiom: aggregation via hierarchical clustering (visible)

Hierarchical Clustering Explorer

Dimensionality reduction
• attribute aggregation
  – derive low-dimensional target space from high-dimensional measured space
  – use when you can’t directly measure what you care about
  – true dimensionality of dataset compressed to be smaller than dimensionality of measurements
• latent factors, hidden variables

Dimensionality vs attribute reduction
• vocab use in field not consistent
• dimensionality
  • attribute reduction: reduce set with filtering
    – includes orthographic projection
  • dimensionality reduction: create smaller set of new dims/attribs
    – typically implies dimensional aggregation, not just filtering
    – vocab projection/mapping

System: Hierarchical Clustering Explorer


Understanding synthetic dimensions

- Specular-Metallic
- Diffuseness-Glossiness

Interacting with dimensionally reduced data


Dimensionality reduction & visualization

- why do people do DR?
  - improve performance of downstream algorithm
  - avoid curse of dimensionality
  - data analysis
- abstract tasks when visualizing DR data
  - dimension-oriented tasks
  - naming synthesized dims, mapping synthesized dims to original dims
  - cluster-oriented tasks
  - verifying clusters, naming clusters, matching clusters and classes

Cluster-oriented tasks

- verifying, naming, matching to classes
- as discernable clusters, clearly discernable clusters, clear match clusters, partial match clusters, no match clusters

Linear dimensionality reduction

- principal components analysis (PCA)
  - first try: PCA (linear)
  - second try: charting (nonlinear DR technique)
  - result: errors fall off sharply after ~45 dimensions
  - result: 104 high-res images of material
  - proc can handle curved rather than linear structure

Nonlinear DR

- many techniques proposed
  - many literatures: visualization, machine learning, optimization, psychology...
  - t-SNE: excellent for clusters
  - but some trickiness remains:
    - http://distill.pub/2016/misread-tsne/

Finding semantics for synthetic dimensions

- looking for meaning in scatterplots
- people inspect images corresponding to points to decide if axis could have meaningful name
- cross-check meaning
- arrows show simulated images (teapots) made from model


Idiom: DOITrees Revisited

- partial match
- clear match
- no match
- some items dynamically filtered out
- some items dynamically aggregated together
- some items shown in detail

VDA with DR example: nonlinear vs linear

- DR for computer graphics reflectance model
  - goal: simulate how light bounces off materials to make realistic pictures
  - computer graphics: BRDF (reflectance)
  - idea: measure what light does with real materials

Dimensional d

Idiom: Dimensionality reduction for documents

- why do people do DR?
  - improve performance of downstream algorithm
  - avoid curse of dimensionality
  - data analysis
- abstract tasks when visualizing DR data
  - dimension-oriented tasks
  - naming synthesized dims, mapping synthesized dims to original dims
  - cluster-oriented tasks
  - verifying clusters, naming clusters, matching clusters and classes

Cluster-oriented tasks

- verifying, naming, matching to classes
- as discernable clusters, clearly discernable clusters, clear match clusters, partial match clusters, no match clusters

Linear dimensionality reduction

- principal components analysis (PCA)
  - first try: PCA (linear)
  - second try: charting (nonlinear DR technique)
  - result: errors fall off sharply after ~45 dimensions
  - result: 104 high-res images of material
  - proc can handle curved rather than linear structure

Nonlinear DR

- many techniques proposed
  - many literatures: visualization, machine learning, optimization, psychology...
  - t-SNE: excellent for clusters
  - but some trickiness remains:
    - http://distill.pub/2016/misread-tsne/

Finding semantics for synthetic dimensions

- looking for meaning in scatterplots
- people inspect images corresponding to points to decide if axis could have meaningful name
- cross-check meaning
- arrows show simulated images (teapots) made from model
- check if those match dimension semantics

Dimensions-oriented tasks

- naming synthesized dims: inspect data represented by lowD points
- abstract location of each point as linear combination of weights for each axis
- mapping synthesized dims to original dims

Linear DR

- first try: PCA (linear)
  - result: errors fall off sharply after ~45 dimensions
  - result: 104 high-res images of material
  - proc can handle curved rather than linear structure

Nonlinear DR

- second try: charting (nonlinear DR technique)
  - result: errors fall off sharply after ~45 dimensions
  - result: 104 high-res images of material
  - proc can handle curved rather than linear structure

Understanding synthetic dimensions

- Specular-Metallic
- Diffuseness-Glossiness

Embed: Focus+Context

- combine information within single view
  - elide
    - selectively filter and aggregate
  - superimpose layer
    - local lens
  - distortion design choices
    - region shape: radial, rectilinear, complex
  - how many regions: one, many
  - region extent: local, global
  - interaction metaphor

Ch 14: Embed
Hierarchical Clustering Explorer
- heatmap, dendrogram
- multiple views


System: VisDB
VisDB Results
- partition into many small regions: dimensions grouped together

VisDB System
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

VisDB Analysis
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

Scagnostics analysis
- scatterplot diagnostics
- scagnostics SPLOM: each point is one original scatterplot

Ch 15: Case Studies

Graph-Theoretic Scagnostics
- scatterplot diagnostics
- scagnostics SPLOM: each point is one original scatterplot

VisDB Results
- partition into many small regions: dimensions grouped together

VisDB Analysis
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

Scagnostics analysis
- scatterplot diagnostics
- scagnostics SPLOM: each point is one original scatterplot

VisDB System
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

VisDB Analysis
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

Ch 15: Case Studies

Graph-Theoretic Scagnostics
- scatterplot diagnostics
- scagnostics SPLOM: each point is one original scatterplot

VisDB Results
- partition into many small regions: dimensions grouped together

VisDB Analysis
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

Scagnostics analysis
- scatterplot diagnostics
- scagnostics SPLOM: each point is one original scatterplot

VisDB System
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

VisDB Analysis
- table: draw pixels sorted, colored by relevance
- group by attribute or partition by attribute into multiple views

Ch 15: Case Studies
InterRing Analysis

**System**
- **What Data**: Tree.
- **Why Tasks**: Selection, rolloff/rollover, hierarchy editing.
- **How Encode**: Radial, space-filling layout. Color by tree structure.
- **How Facet**: Linked coloring and highlighting.
- **How Reduce**: Embed, distort, multiple face.
- **Scale**: Nodes: hundreds, if labeled; thousands if dense. Levels in tree: dozens.

**Using Space: Constellation**
- **Data**: multi-level network
  - node: word
  - link: words used in same dictionary definition
  - subgraph: for each definition
  - visual encoding:
    - link connection marks between words
    - link containment marks to indicate subgraph
    - encode probability with hexa spatial position
    - encode source in terms of the word's spatial position
    - aesthetic layout

**Using Space: PivotGraph**
- **Data**: derived rollup network
  - visual encoding:
    - link connection marks between words
    - link containment marks to indicate subgraph
    - encode probability with hexa spatial position
  - edge crossings
  - cannot easily minimize instances, since position constrained by spatial encoding
  - must preserve graph-theoretic properties
  - violation: single dataset, many visualizations
  - violation: many datasets, same visualization

**Visual Exploration of Multivariate Graphs**
- **Data**: derived rollup network
  - visual encoding:
    - link connection marks between words
    - link containment marks to indicate subgraph
    - encode probability with hexa spatial position
  - edge crossings
  - cannot easily minimize instances, since position constrained by spatial encoding
  - must preserve graph-theoretic properties
  - violation: single dataset, many visualizations
  - violation: many datasets, same visualization

**Visual Exploration of Multivariate Graphs, Martin Wattenberg, CHI 2006.**

**Topological Fisheye Views**
- **Data**: multi-level network
  - node: word
  - link: words used in same dictionary definition
  - subgraph: for each definition
  - visual encoding:
    - hybrid view made from cut through several hierarchy levels

**Coarsening requirements**
- uniform cluster/metadata size
- match coarse and fine layout geometries
- scalable

**Coarsening strategy**
- must preserve graph-theoretic properties
- use both topology and geometry
- toposological distance (hops away)
- geometric distance - but not just proximity alone!
- can compact neighborhoods could create new cycles
- derived data: proximity graph

**Visual Exploration of Large Graphs, Martin Wattenberg, CHI 2006.**
Biomechanical motion design study
- large DB of 3D motion data
  - pigs chewing: high-speed motion at joints, 500 FPS w/ sub-mm accuracy
- goal: to understand functional morphology: relationship between 3D shape of bones and their function

Hybrid graph creation
- cut through coarsening hierarchy to get active nodes
  - animated transitions between states

Final distortion
- geometric distortion for uniform density
  - (coloroded by hierarchy depth just to illustrate algorithm)
- compare to original
  - (compared to simple topologically aware fish-eye distortion)

Example Presentation: Biomechanical Motion

Presentation expectations
- 20 minute time slots for presentations
  - slides required
    - if you're using your laptop, send me by 12pm
  - three goals up to you whether sequential or interleaved
    - explain core technical content to audience
    - analysis with doing what/how framework
    - do scale analysis of data for this system in specific, not for technique in general

Analysis & critique
- paper type dependent
  - required for design studies and technique papers
  - some possible for algorithm papers
  - but more emphasis on presenting algorithm clearly
  - they should play with their own papers
  - but can discuss study design and statistical analysis methods

- please distinguish: their analysis (future work, limitations) from your own thoughts/critiques
  - good to present both

Beyond paper itself
- check for author paper page
  - may have video
  - may have talk slides you could borrow as a base
  - do acknowledge if so!

Slides
- do include both text and images
  - font must be readable from back of room
  - bullet style not sentences
  - use laser pointer judiciously

Technical talks advice
- How To Give An Academic Talk
  - Paul N. Edwards

- How To Give a Great Research Talk
  - Simon L. Peyton Jones, John Hughes, and John Laurie:

- How To Present A Paper
  - Leslie Lamport

- Things I Hope Not To See or Hear at SIGGRAPH
  - Jim Blev

Scientific Presentation Planning
- Example Presentation:
  - face audience, not screen
  - pro tip: your screen left/right matches audience left/right in this configuration
  - hard to follow if parts only visible partly

- Images alone often hard to follow
  - images do not speak for themselves, you must walk us through them
  - text bullets to walk us through your highest-level points
  - hard to follow if parts only visible partly
  - judgement call on text/image ratio, avoid extremes

Style
- figures from paper
  - good idea to use figures from paper, especially screenshots
  - judgement call about how many

- new images
  - you might make new diagrams
  - you might grab other images, especially for background or if comparing to prev work
  - avoid random clip art

- images alone often hard to follow
  - images do not speak for themselves, you must walk us through them

Biomechanical motion design study
- large DB of 3D motion data
  - high-speed motion at joints, 500 FPS w/ sub-mm accuracy
- domain tasks
  - functional morphology: relationship between 3D shape of bones and their function
- abstract tasks
  - trends & anomalies across collection of time-varying spatial data

- pioneering design study integrating infovis+scivis techniques
- let's start with video showing system in action

Candidate pairs: neighbors in original and proximity graph
- derived data: traces/streamers
  - derived data: 3D motion tracers
  - generated x/y data over time
  - streamers
  - shown in 3D views directly
  - populates 2D plots

Multiple linked spatial & non-spatial views
- data: 3D spatial, multiple attributes (cyclic)
  - encode 3D spatial, parallel coordinates, 2D line (ey) plots
- facet: few large multifield views, many small multiples (~100)
  - encode color by trial for window background

- view coordination
  - line in percolated view frame in small multi

3D+2D
- change
  - 3D navigation
  - rotate/translate/zoom
  - filter
  - zoom to small subset of time

Example Presentation: Interactive Coordinated Multiple-View Visualization of Biomechanical Motion Data
Small multiples for overview

• facet: small multiples for overview
  – aggressive/ambitious, 100+ views
• encode: color code window bg by trial
  – full/partial skull
  – streamers
  – simple enough to be usable at low information density

Analysis summary

• what: data
  – 3D spatial, multiple attribs (cyclic)
• what: derived
  – 3D motion traces
  – 3D surface interaction patterns
• facet: linked navigation
  – superimposed overlays in 3D view
• encoding: color coding

Derived data: surface interactions

• what: derived
  – 3D surface interaction patterns
• facet: superimposed overlays in 3D view
• encoding: colored by vertical distance separating teeth (derived surface interactions)

Side by side views demonstrating tooth slide

• facet: linked navigation w/ same 3D viewpoint for all
• encode: colored by vertical distance separating teeth (derived surface interactions)
  – also 3D instantaneous helical axis showing motion of mandible relative to skull

Cluster detection

• identify clusters of motion cycles
  – from combo: 2D xy plots & parcoords
  – show motion itself in 3D view
• facet: superimposed layers
  – foreground/background layers in parcoord view itself

Critique

• many strengths
  – carefully designed with well justified design choices
  – explicitly followed mantra “overview first, zoom and filter, then details-on-demand”
  – sophisticated view coordination
  – tradeoff between strengths of small multiples and overlays, use both
  – enhanced by difficulties of animation for trend analysis
  – derived data tracing paths
• weaknesses/limitations
  – (older paper feels less novel, but must consider context of what was new)

Next time

• deadlines
  – meetings due by Thu Nov 2, 5pm
  – proposals due by Mon Nov 6, 10pm
• next week
  – presentations 1
  – guest lecture from Steve Franconeri