Information Visualization
Aggregate & Filter 2

Tamar Munzner
Department of Computer Science
University of British Columbia

Lect 19, 17 Mar 2020

News
• Online lectures and office hours start today, using Zoom: https://zoom.us/j/9016202871
• Lecture mode
  – Plan: I’ll livestream with video + audio + screenshare, will also try recording.
  – You’ll be able to just join the session
  – Please connect audio-only, no video, to avoid congestion
  – You’ll be auto-muted. If you have a question use the Show Hand (click on Participants, button is at the bottom of the popup window), I’ll unmute you myself
• Office hours mode
  – Please do connect with video if possible, in addition to audio
  – I’ll use the Waiting Room feature, where I will individually allow you in
  – If I’m already talking to somebody else I’ll briefly let you know, then put you back in VR until it’s your turn.
• P1-P3 marks
  • increasingly bimodal
  • P3 marks released
  • bimodal distribution

Spatial aggregation
• MAUP: Modifiable Areal Unit Problem
  – changing boundaries of cartographic regions can yield dramatically different results
  – scale effects

Gerrymandering: MAUP for political gain
Example: Gerrymandering in PA

Idiom: GrouseFlocks
• data: compound graphs
  – network
  – derived or interactively chosen
  – cluster hierarchy step it

Idiom: Hierarchical parallel coordinates
• dynamic item aggregation
• derived data: hierarchical clustering
• encoding:

Schedule shift
• Nothing due this Wed
• M2 & M3 on schedule
  – M2 due Wed Mar 25
  – M3 due Wed Apr 8
• Combined F5/F6
  – will go out Thu Mar 26, due Wed Apr 1

Foundations F1-F4


Cluster
• classification of items into similar bins
  – based on similarity measure
    • Euclidean distance, Pearson correlation
  – partitioning algorithms
  – divisive into set of bins
  – K-means (it is manual or automatically
  – hierarchical algorithms
  • produce “similarity tree” (dendrogram): cluster hierarchy
  – agglomerative clustering: start w/ each node as own cluster, then iteratively merge
  – cluster hierarchy: derived data used w/ many dynamic aggregation idioms
  – cluster more homogeneous than whole dataset
  • statistical measures & distribution more meaningful

Idiom: Hierarchical Clustering (visible)
System: Hierarchical Clustering (bottom)

Idiom: aggregation via hierarchical clustering (visible)

News
• Labs will be Zoom + Canvas scheduling
  – different Zoom URL for each TA, stay tuned
  – you can sign up for reserved slots in advance, or check for availability on the fly
  – more details soon
• Final exam plan still TBD
  – but will not be in person
  – you are free to leave campus when you want (but are not required to do so)

Nonlinear DR

- second try: charting (nonlinear DR technique)
  - scree plot suggests 10-15 dims
  - note: dim estimate depends on technique used!
  - dimension-oriented tasks
    - naming synthesized dims; mapping synthesized dims to original dims
    - cluster-oriented tasks
      - verifying clusters; naming clusters; matching clusters and classes

Nonlinear dimensionality reduction

- pro: can handle curved rather than linear structure
- cons: lose all ties to original dims/attrs
- many techniques proposed:
  - many literatures: visualization, machine learning, optimization, psychology, ...
  - techniques: t-SNE, MDS (multidimensional scaling), charting, isomaps, LLE, ...
  - t-SNE: excellent for clusters
  - but some tricks remain:
    - MDS: confusingly, entire family of techniques, both linear and nonlinear
    - minimize stress or strain metrics
    - early formulations equivalent to PCA

Linear DR

- first try: PCA (linear)
  - result: errors fall off sharply after ~45 dimensions
  - problem: physically impossible intermediate points when simulating new materials
  - speculative highlights could have holes!

Dimensionality reduction & visualization

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

Finding semantics for synthetic dimensions

- look for meaning in scatterplots
  - screen plots w/ nearest neighbors
  - points represent real-world images (spheres)
    - points represent real-world images (spheres)
  - people inspect images corresponding to points to decide if axis could have meaningful name
  - cross-check meaning
    - arrows show simulated images (isomaps) made from model
    - check if those match dimensional semantics

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)
  - goal: simulate how light bounces off materials to make realistic pictures
  - computer graphics: BRDF (reflectance)
- capturing & using material reflectance
  - result: 104 high-res images of material
    - each image 4M pixels
  - goal: capture most of variance with minimal error
  - computer graphics: BRDF (reflectance)
  - goal: simulate how light bounces off materials to make realistic pictures
  - computer graphics: BRDF (reflectance)
  - goal: capture most of variance with minimal error
  - computer graphics: BRDF (reflectance)

Understanding synthetic dimensions

- Specular-Metallic
- Diffuse-Glossy
- transparent

Dimensionality reduction

- attribute aggregation
  - derive low-dimensional target space from high-dimensional measured space
  - capture most of variance with minimal error
  - true dimensionality of dataset conjectured to be smaller than dimensionality of measurements
  - bias factors, hidden variables

Linear dimensionality reduction

- principal components analysis (PCA)
  - finding axes: first with most variance, second with next most, ...
  - describe location of each point as linear combination of weights for each axis
  - mapping synthesized dims to original dims

Nonlinear dimensionality reduction

- use to capture most of variance with minimal error
- techniques: t-SNE, MDS (multidimensional scaling), charting, isomaps, LLE, ...
- t-SNE: excellent for clusters
- but some tricks remain:
  - MDS: confusingly, entire family of techniques, both linear and nonlinear
  - minimize stress or strain metrics
  - early formulations equivalent to PCA

Finding semantics for synthetic dimensions

- look for meaning in scatterplots
  - screen plots w/ nearest neighbors
  - points represent real-world images (spheres)
  - people inspect images corresponding to points to decide if axis could have meaningful name
  - cross-check meaning
    - arrows show simulated images (isomaps) made from model
    - check if those match dimensional semantics

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)
  - goal: simulate how light bounces off materials to make realistic pictures
  - computer graphics: BRDF (reflectance)
  - goal: capture most of variance with minimal error
  - computer graphics: BRDF (reflectance)
  - goal: simulate how light bounces off materials to make realistic pictures
  - computer graphics: BRDF (reflectance)
  - goal: capture most of variance with minimal error
  - computer graphics: BRDF (reflectance)

Linear DR

- first try: PCA (linear)
  - result: errors fall off sharply after ~45 dimensions
  - problem: physically impossible intermediate points when simulating new materials
  - speculative highlights could have holes!
Idiom: DoITrees Revisited

- 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
- distortion costs and benefits
  - benefits
    - combine focus and context information in single view
  - costs
    - length comparisons impaired
    - network/tree topology comparisons unaffected
    - effects of distortion unclear if original structure unfamiliar
  - effects of distortion unclear if original structure unfamiliar
  - object constancy/tracking impaired

Idiom: Fisheye Lens

- distort geometry
  - shape: radial
  - focus: single extent
  - extent: local
  - metaphor: draggable lens

System: TreeJuxtaposer


Credits

- Visualization Analysis and Design (Ch 13, 14)
- Alex Lex & Miriah Meyer, http://dataviscourse.net/