Problem-Driven Interactive Visualization for Imperfect Models

Tamar Munzner
Department of Computer Science
University of British Columbia

@tammamunzner

http://www.cs.ubc.ca/~tmm/talks.html#huawei22

Problem-Driven Interactive Visualization for Imperfect Models
Tamara Munzner
Department of Computer Science
University of British Columbia
Huawei Vancouver
Jan 19 2022, virtual
@tamaramunzner

Visualization (vis) defined & motivated

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, people can see a lot in the details
- confirm expected and find unanticipated patterns
- issues with statistical models
- sensitivity analysis for parameters

Why is validation difficult?

- Unpacking data visually: From rollup to drilldown
- human in the loop needs details about data
- is the user looking at it?
- is the user looking at it?
- is the user looking at it?
- is the user looking at it?
- why is it shown?
- task abstraction
- data abstraction
- how is it shown?
- visual encoding idiom: how to draw
- interaction idiom: how to manipulate
- algorithm
- efficient computation

Nested model: Four levels of visualization concerns

- domain situation
  - who are the target users?

- abstraction
  - translate from specifics of domain to vocabulary of vis
  - what is shown?
  - data abstraction
  - data abstraction
  - often don't just draw what you're given; transform to new form

- task abstraction
  - why is the user looking at it?
  - task abstraction

- idiom
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

Outline

- methodology of problem-driven visualization research
- two case studies of visualizing imperfect models
  - NLP for temporal data
  - ML with graph neural networks
- brief overview of other problem-driven projects

Why is validation difficult?

- different ways to get it wrong at each level

Unpacking data visually: From rollup to drilldown

Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- summaries lose information, people can see a lot in the details
- confirm expected and find unanticipated patterns
- issues with statistical models
- sensitivity analysis for parameters

Why is validation difficult?

- different ways to get it wrong at each level

Updated: 2022-01-19

More at:

http://www.cs.ubc.ca/~tmm/talks.html#huawei22

A Nested Model for Visualization Design and Validation


A Nested Model for Visualization Design and Validation.


Datasets with Varied Appearance and Identical Statistics

Matejka & Fitzmaurice

Datasets with Varied Appearance and Identical Statistics

Matejka & Fitzmaurice


A Multi-Level Typology of Abstract Visualization Tasks

Brehmer and Munzner. IEEE TVCG

[A Multi-Level Typology of Abstract Visualization Tasks

Brehmer and Munzner. IEEE TVCG


Visual encoding/interaction idiom

Task abstraction

Abstraction

Domain situation

- who are the target users?

- idiom
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- algorithm
  - efficient computation

A Nested Model for Visualization Design and Validation


A Nested Model for Visualization Design and Validation.


Visual encoding/interaction idiom

Task abstraction

Abstraction

Domain situation

- who are the target users?

- idiom
  - how is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate

- algorithm
  - efficient computation

A Nested Model for Visualization Design and Validation


A Nested Model for Visualization Design and Validation.


Visual encoding/interaction idiom

Task abstraction

Abstraction

Domain situation

- who are the target users?
Why is validation difficult?
• different ways to get it wrong at each level
  
  Domain situation
  - You misunderstood their needs
  - Design: observed target users using existing tools
  - Data/task abstraction: target users using existing tools
  - Algorithm: target users after deployment (dark study)
  - Anthropology/ethnography: observed target users using existing tools
  
  Validation solution: use methods from appropriate fields at each level
  - Domain situation
  - Data/task abstraction
  - Algorithm
  - Anthropology/ethnography
  
  computer science
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
  
  Why is validation difficult?
  - different ways to get it wrong at each level
  
  Domain situation
  - You misunderstood their needs
  - Design: observed target users using existing tools
  - Data/task abstraction: target users using existing tools
  - Algorithm: target users after deployment (dark study)
  - Anthropology/ethnography: observed target users using existing tools
  
  Validation solution: use methods from appropriate fields at each level
  - Domain situation
  - Data/task abstraction
  - Algorithm
  - Anthropology/ethnography
  
  computer science
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)

Visualization: Angles of attack

Validation solution: use methods from appropriate fields at each level
  
  computer science
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)

Why is validation difficult?
• different ways to get it wrong at each level
  
  Domain situation
  - You misunderstood their needs
  - Design: observed target users using existing tools
  - Data/task abstraction: target users using existing tools
  - Algorithm: target users after deployment (dark study)
  - Anthropology/ethnography: observed target users using existing tools
  
  Validation solution: use methods from appropriate fields at each level
  - Domain situation
  - Data/task abstraction
  - Algorithm
  - Anthropology/ethnography
  
  computer science
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)

Why is validation difficult?
• different ways to get it wrong at each level
  
  Domain situation
  - You misunderstood their needs
  - Design: observed target users using existing tools
  - Data/task abstraction: target users using existing tools
  - Algorithm: target users after deployment (dark study)
  - Anthropology/ethnography: observed target users using existing tools
  
  Validation solution: use methods from appropriate fields at each level
  - Domain situation
  - Data/task abstraction
  - Algorithm
  - Anthropology/ethnography
  
  computer science
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
  - Algorithm: Resource management (monitoring)
“A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts...”
Design studies & user-centered design

- user-centered design: well-known HCI methodology
  - iterative refinement & deployment
  - evaluation through case studies & field studies

what's specific to visualization?
- discovering task and data abstractions
- designing visual encoding & interaction idioms that map to abstractions

Two case studies of visualizing imperfect models

- NLP for temporal data
- ML for graph data

ML for graph data
- link prediction
- node classification
- examples: product recommendation, protein interactions

NLP for temporal data
- cool use case: eureka moment
  - success: enable what was impossible before
  - vis tools for new insights & discoveries
  - workhorse use case: workflow speedup
  - sometimes enables the previously infeasible

The importance of being brisk
- two other use cases nudge towards exploration
- comparison between multiple timelines
- speculative browsing

Graph neural network (GNN)
- machine learning (ML) models for graphs
  - like CNN for images
  - like Transformer for text
- many real-world graph-related applications
  - node classification
  - example: fraud detection, disease classification
  - link prediction
  - example: product recommendation, protein interactions
Problem-driven visualization for imperfect models

- problem-driven methodology
  - translate domain problems into abstractions
  - before visual encoding idioms & algorithms
  - avoid collaboration pitfalls
  - understand roles, ensure aligned incentives

- interactive visualization supporting human-in-the-loop judgments about models
  - two cases: different data types

- overview: other problem-driven projects

More information

- this talk
  [http://www.cs.ubc.ca/~tmm/talks.html#huawei22](http://www.cs.ubc.ca/~tmm/talks.html#huawei22)
- papers, videos, software, talks, courses
  [http://www.cs.ubc.ca/~tmm](http://www.cs.ubc.ca/~tmm)
- theoretical foundations: book
  [+ tutorial/course lecture slides]