Storytelling (Narrative Visualization)

Maneesh Agrawala
Jessica Hullman

CS 294-10: Visualization
Fall 2014

Last Week: Text Visualization
Doc. Similarity & Clustering

In vector model, compute distance among docs
For TF.IDF, typically cosine distance
Similarity measure can be used to cluster

ThemeRiver (Havre et al 99)
Parallel Tag Clouds (Collins et al 09)

Topic modeling

Topic modeling approaches
Assume documents are a mixture of topics
Topics are (roughly) a set of co-occurring terms
Latent Semantic Analysis (LSA): reduce term matrix
Latent Dirichlet Allocation (LDA): statistical model
How to visualize results? (Blei 2012)

Typical display of topics is by their most frequent words.

How can we best display ....

- a document with a topic model?
- connections between documents via hidden topic proportions?
- a whole corpus and its inferred topic structure?
Plotting topics

Blei 2012

Termite (Chuang et al. 2012)
Interpretable and trustworthy

Chuang et al. 2012
Takeaways

Show (or provide access to) source text
Let readers assess model
Let readers use visualization as index into documents

Find meaningful abstractions for grouping docs
Are clusters interpretable?

Where possible use text to represent text… but which terms are the most descriptive?

---

Natural Science and Engineering
(Ellingham 1948)

Manual analysis, layout, and labeling
Treemap
Central disciplines
Sub-disciplines
Topics

H. J. T. Ellingham 1948
Royal Society Scientific Information Conference
Geography of Science (Small et al. 1985)

Citation analysis
Hierarchical clustering
Levels of details
Edge labeling

Henry Small, Eugene Garfield 1985

The geography of science, Journal of Information Science
Melanoma Research (Boyack et al. 2004)

Papers, genes, proteins
Landmark labeling (manual)

Kevin Boyack, Ketan Mane, Katy Börner 2004

Mapping medline papers, genes, and proteins related to Melanoma research, IV2004 Conference
Scholarly Impact (Rosvall et al. 2007)

Map of random walks on complex network reveal community structure, PNAS

Text visualization collection

http://textvis.lnu.se/
Storytelling
Crimean War Deaths [Nightingale 1858]
**Janet L. Yellen, on the Economy’s Twists and Turns**

**Inflation and unemployment**

The Federal Reserve is said to have a “dual mandate”-keeping inflation in check and the unemployment rate low. These measures, which tend to change cyclically and in concert with each other, are charted for every year since the Great Depression.

In speeches and in meetings, Ms. Yellen, the nominee for the next Fed leader, has commented on the Fed’s actions during significant periods, providing a window into her views and priorities.

---

**Narrative Visualization**
Storytelling in visualization

Narrative visualization
• Genres, Design features

Construction
• Design process
• Automated generation
• Design tools for narrative visualizations

Value of storytelling (Gershon and Page 2001)

A way of structuring information
• Easier to understand than lists
• Uncertainty, conflict, resolution
• Text and visuals can be complementary

Data exploration Communication
Storytelling techniques

- build in picture
- continuity
- animate events
- Increasing attention
- redundancy

Narrative Visualization (Segel and Heer 2010)

Studied 58 examples
- News media, blogs, instructional videos, research

Characterizing the design space
Narrative Visualization (Segel and Heer 2010)

Design space dimensions: 1. Genre
2. Visual narrative tactics

Highlighting

Transition guidance

3. Narrative structure tactics

Ordering
Messaging
Interactivity
Author vs. reader driven stories

Reader driven:
- Multiple orderings
- Less messaging
- More open interactivity

Author driven:
- Prescribed ordering
- Stronger messaging
- Limited interactivity

'How the Recession Changed Us'

<table>
<thead>
<tr>
<th>DAILY CONSUMER SPENDING</th>
<th>FEATURE-LENGTH MOVIES released in 3-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>October 2008</td>
<td>October 2010</td>
</tr>
<tr>
<td>$81</td>
<td>$82</td>
</tr>
<tr>
<td>$25.9 billion</td>
<td>$23.9 billion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>U.S. BOOK SALES</th>
<th>AVERAGE MINUTES SPENT PLAYING A GAME or USING A COMPUTER for leisure per WEEKEND DAY, 15-to-19-year-olds</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>2009</td>
</tr>
<tr>
<td>$25.9 billion</td>
<td>12</td>
</tr>
<tr>
<td>$23.9 billion</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PERCENTAGE INCREASE in NYT ARTICLES mentioning the WORD “uncertainty,” 2007 vs. 2010*</th>
</tr>
</thead>
<tbody>
<tr>
<td>44%</td>
</tr>
</tbody>
</table>

| TAX COLLECTED by the TREASURY on sales of FIREARMS and AMMUNITION |
|================================================================||
| 2007                       | 2009                           |
| $18.5 billion              | $14.8 billion                 |

*through Dec. 3
Visualization rhetoric

Rhetoric / framing: How an interpretation arises from a representation, plus individual, social factors

Framing effects: small changes in presentation of an issue result in significant changes in opinion

Method
• 51 professional produced narrative visualizations
• NYT, BBC, Economist, local news, political outlets
• Iterative qualitative coding, seeded scheme with semiotics, persuasion concepts
**Design process**

An editorial process characterized by rhetorical decisions at various points.

- **Filter, transform**
  - Data

- **visualize**

- **select**

- **annotate**

- **design interactions**

**Taxonomy of Rhetorical Strategies**

**Rhetorical Categories**
- Information Access
- Provenance
- Mapping
- Procedural
- Linguistic

**Editorial Layers**
- Data
- Visual Representation
- Annotation
- Interactivity

- Omission (variable selection)
- Metonymy (aggregation and categorization)
- Visual metaphor
- Semantic encodings
- Citing sources, methodology
- Exceptions, corrections
- Acknowledging uncertainty
- Rhetorical question, “scare quotes”
- Labeling choices
- Analogy, simile, double entendre
**Enthymeme** (information access)

- DAILY CONSUMER SPENDING
  - October 2008: $\$\$\$\$\$\$
  - October 2010: $\$\$\$\$\$

- FEATURE-LENGTH MOVIES released in 3-D
  - 2007: 4
  - 2010: 23

- U.S. BOOK SALES
  - 2007: $25.9$ billion
  - 2009: $23.9$ billion

- PERCENTAGE INCREASE in NYT ARTICLES mentioning the WORD "uncertainty," 2009 vs. 2010:
  - $44\%$

- TAX COLLECTED by the TREASURY on sales of FIREARMS and AMMUNITION

- AVERAGE MINUTES SPENT READING per WEEKEND DAY, 12- to 19-year-olds
  - 2007: 18
  - 2009: 5

- AVERAGE MINUTES SPENT PLAYING A GAME or USING A COMPUTER for leisure per WEEKEND DAY, 12- to 19-year-olds
  - 2007: 24
  - 2009: 11

**Visual metaphor** (mapping rhetoric)

- Diagrams illustrating organizational charts of the House American Youth Plan

- Flowchart showing connections and relationships in a complex system
Labeling, sourcing (linguistic, provenance)

Take-aways

Narrative visualizations blend communication / exploratory techniques

Specific strategies can be identified by studying what professionals make

Messaging, metaphor, other suggestive strategies have a powerful impact on interpretation
Designing narrative visualizations

“require skills like those familiar to movie directors, beyond a technical expert’s knowledge of computer science and engineering….”

*Gershon and Page 2001*
Lyra (Satyanarayan and Heer 2014)

Tableau Story Points

Austin's Teacher Turnover Problem

Many Austin schools are still struggling to hold onto teachers, despite a multimillion-dollar effort, asked Rees, to stem turnover in troubled schools since 2010.

Austin's teacher turnover is above the national average. Teacher turnover is a city-wide problem, and the problem is particularly acute at some schools. And city-wide, the problem isn't getting better.

Austin's teacher turnover rate by year. Turnover was above the national average for most years.
AI-generated stories

Artificial intelligence + narratology

• ‘Tell me a story’ (Schank 1990)

Narrative Science

Visualizations with news articles
Annotations tell the story

Can we produce an automated solution?

Existing automated solution

Can we identify criteria designers use to add context?

Examined annotation content, placement (function) in 136 news visualizations

Add relevant info

• Data
• Mapping
• Context

Observe salient values

Stock data

Company performance
**Contextifier system**

Input: news article about a company

- Yahoo Decides to Friend Facebook
- Closing price
- Volume sold

(process text)

Query

Article Corpus

Generate features,
Select annotations

(obtain stock data)

**Relevancy feature**

Kullback-Leibler (KL) divergence

\[ D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \]
**Article volume**

Prioritize dates with unusually high numbers of company articles.

**Visual saliency**

Visual saliency proxy via simple analyses on stock series.

$\max(t), \min(t) \quad t_i - t_{i-1}$
3*relevancy + saliency + 0.5 volume

Extension to geovisualization

The New York Times

U.S.

Obesity, Diabetes Reach Epidemic Levels

Obesity and diabetes have both reached epidemic levels in the United States, placing health authorities who have been warning for years on the spot.

By the end of 2000, 103 million American adults were obese and 15.8 million were diabetic, according to a government study released today by the Centers for Disease Control and Prevention. Obesity and diabetes have increased in every category—age, rural and urban, white and black, rich and poor. Yet the U.S. health care system is overburdened for the world epidemic.

All told, 27 million people suffer from obesity, a number that is expected to rise to 40 million by 2010.

From 1980 to 1998, the rate of obesity climbed to 20.5 percent of adults from 10.5 percent, and the rate of diabetes rose to 7.1 percent from 5.1 percent.

The survey, based on a telephone interview of 100,000, shows that the rate of diabetes is not limited to those who are overweight. Obesity is more likely among people who are thin, but diabetes is more common among those 60 and older. The rate of diabetes is also much higher among women.

The latest survey also shows that obesity is rising faster among those 65 and older, and that the rate of diabetes is rising faster among those 65 and older, and that the rate of diabetes is rising faster among those 65 and older, and that the rate of diabetes is rising faster among those 65 and older, and that the rate of diabetes is rising faster among those 65 and older, and that the rate of diabetes is rising faster among those 65 and older, and that the rate of diabetes is rising faster among those 65 and older.

In the United States, the rate of obesity is higher among African-Americans, Hispanics, and Native Americans, and the rate of diabetes is higher among African-Americans, Hispanics, and Native Americans.

The cost of diabetes to the nation includes billions of dollars in cost for the heart disease, high blood pressure, kidney failure, arthritis, and blindness that are the long-term effects of obesity and diabetes.
Generalizing to geovisualization

NewsViews system tools

Fort Worth High School Graduate Captures Aaja Title

Educational Attainment 25 Years And Over High School Graduate Or Higher (percent)

<table>
<thead>
<tr>
<th>0 - 66</th>
<th>66 - 73.3</th>
<th>73.3 - 78.4</th>
<th>78.4 - 82.8</th>
<th>82.8 - 86.7</th>
<th>86.7 - 90.4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NewsViews system tools

In 2011, Wyoming has the highest
percent.

Women Evaluation in Education Achievement

The NewsViews system is being used to provide
information on educational achievement for women.

Fort Worth High School Graduate Captures Aaja Title

In 2011, Wyoming has the highest percent.

Women Evaluation in Education Achievement

The NewsViews system is being used to provide
information on educational achievement for women.

NewsViews system tools

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.

Recent articles on the NewsViews website include
information on educational achievement for women.
Other approaches

Existing systems are fully automated or require a designer with some expertise

What if designer drives, but system makes suggestions?

Requirements:
- Model for effectiveness of narrative features
- Identify where to fit into the design process

Effectiveness of stories

Genre, presentation medium will affect learning experience of user (Gershon and Page 2001)

Systematic study of narrative
- Macro- and micro-structure
- Presentation order impacts conclusion (Kahneman and Tversky 1972, Pennington & Hastie 1992)
- Canonical order (Thorndyke 1977)
Sequenced visualizations

Wrong order = wrong story -
define context / storytelling goal

Support sequencing in design process

filter
visualize
select
annotate
design interaction, order

Study 1 – Identify sequence principles

Expert practice
400+ transitions in explicitly ordered visualizations

Patterns, principles
Transition principles

Design support
Single transition principles

Dimension walk

“Fundamental” types
Transition as single change to data schema
• Transition cost
Study 2 – Evaluate effect of principles

Evaluated principles with online audiences:
- Transition cost
  - Low cost is easier to understand
- Type differences
  - Temporal > Measure/Dimension > Hierarchy

Expert practice
Patterns, principles
Connect to reception
Design support

Study 2: multiple transitions – parallelism

Can improve recall and ease of understanding transitions but decrease visual comparisons.