Visual Analytics

CS 4460 – Intro. to Information Visualization
November 20, 2014
John Stasko

Agenda

• Overview of what the term means and how it relates to information visualization
• Some example VA research projects
• Specific example, Jigsaw, helping investigative analysis
Visual Analytics

- A new term for something that is familiar to all of us

- Informal description:
  - Using visual representations to help make decisions
  - Sounds like infovis, no?
Before there was VA

- Growing concern from some that infovis was straying from practical, real world analysis problems
- Infovis typically not applied to massive data sets
- Infovis “competes” with other computational approaches to data analysis
  - Statistics, data mining, machine learning

Important Paper

- Shneiderman suggests combining computational analysis approaches such as data mining with infovis – Discovery tools
  - Too often viewed as competitors in past
  - Instead, can complement each other
- Each has something valuable to contribute

Shneiderman
Information Visualization ’02
Contrasts

- Data mining, machine learning
  - Handle larger data well
  - Better for concrete questions and hypotheses

- Data visualization
  - Enables human judgment and decision making
  - Better for exploratory scenarios

Further Questions

- Are information visualizations helping with exploratory analysis enough?
- Are they attempting to accomplish the right goals?
Another Important Paper

• Information visualization systems inadequately supported decision making:
  – Limited Affordances
  – Predetermined Representations
  – Decline of Determinism in Decision-Making
• “Representational primacy” versus “Analytic primacy”
  – Telling the truth about your data versus providing analytically useful visualizations

Task Level

• Don’t just help “low-level” tasks
  – Find, filter, correlate, etc.

• Facilitate analytical thinking
  – Complex decision-making, especially under uncertainty
  – Learning a domain
  – Identifying the nature of trends
  – Predicting the future
Analytic Gaps

- Analytic gaps — “obstacles faced by visualizations in facilitating higher-level analytic tasks, such as decision making and learning.”
  - Worldview Gap
  - Rationale Gap

Knowledge Precepts

- For narrowing these gaps
  - Worldview-Based Precepts (“Did we show the right thing to the user?”)
    - Determine Domain Parameters
    - Expose Multivariate Explanation
    - Facilitate Hypothesis Testing
  - Rationale-Based Precepts (“Will the user believe what they see?”)
    - Expose Uncertainty
    - Concretize Relationships
    - Expose Cause and Effect
More Motivation

• Increasing occurrences of situations and areas with large data needing better analysis
  – DNA, microarrays
  – 9/11 security
  – Business intelligence

Articulating the Motivation

[Video]

http://videotheque.inria.fr/videotheque/doc/635
History

- 2003-04 Jim Thomas of PNNL, together with colleagues, develops notion of visual analytics
- Holds workshops at PNNL and at InfoVis '04 to help define a research agenda
- Agenda is formalized in book *Illuminating the Path*, shown on next slide

Visual Analytics Definition

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.

People use visual analytics tools and techniques to

- Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, and understandable assessments
- Communicate assessment effectively for action.

“The beginning of knowledge is the discovery of something we do not understand.”
~Frank Herbert (1920 - 1986)
Visual Analytics

- Not really an “area” per se
  - More of an “umbrella” notion
- Combines multiple areas or disciplines

- Ultimately about using data to improve our knowledge and help make decisions

Main Components

- Interactive visualization
- Analytical reasoning
- Computational analysis
Alternate Definition

- Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets.

Keim et al, chapter in *Information Visualization: Human-Centered Issues and Perspectives*, 2008

Synergy

- Combine strengths of both human and electronic data processing
  - Gives a semi-automated analytical process
  - Use strengths from each

From Keim
InfoVis Comparison

- Clearly much overlap
- Perhaps fair to say that infovis hasn’t always focused on analysis tasks so much and that it doesn’t always include advanced data analysis algorithms
  - Not a criticism, just not focus
  - InfoVis has a more narrow scope
  - (Some of us actually do believe that infovis has/should include those topics)

Academic Context

- Visual Analytics ~2005
- Information Visualization ~1990
- Artsy casual infovis, etc.
- Pure analytical reasoning
  - Computational analysis

My interpretation
Visual Analytics

- Encompassing, integrated approach to data analysis
  - Use computational algorithms where helpful
  - Use human-directed visual exploration where helpful
  - Not just “Apply A, then apply B” though
  - Integrate the two tightly

Domain Roots

- Dept. of Homeland Security supported founding VA research
- Area has thus been connected with security, intelligence, law enforcement
- Should be domain-independent, however, as other areas need VA too
  - Business, science, biology, legal, etc.
VA-related Research Topics

- Visualization
  - InfoVis, SciVis, GIS
- Data management
  - Databases, information retrieval, natural language
- Data Analysis
  - Knowledge discovery, data mining, statistics
- Cognitive Science
  - Analytical reasoning, decision-making, perception
- Human-computer interaction
  - User interfaces, design, usability, evaluation

Research Agenda

- Available at http://nvac.pnl.gov/ in PDF form
- At IEEE Press in book form
- Special thanks to IEEE Technical Committee on Visualization and Graphics
Overview of the R&D Agenda

- Challenges
- Science of Analytical Reasoning
- Science of Visual Representations and Interactions
- Data Representations and Transformations
- Production, Presentation, and Dissemination
- Moving Research Into Practice
- Positioning for an Enduring Success

More History

- European Union became very active in visual analytics area
  - VisMaster project
Vision of the Future

- PNNL Precision Info Environments (PIE) video
- Emergency response scenario

Projects

- Let’s look at some recent research projects in this area
sVisit

Meghdadi & Irani
TVCG (VAST) '13

PivotSlice

Zhao, et al
TVCG (VAST) '13
DecisionFlow

Gotz & Stavropoulos
TVCG (VAST) '14

FluxFlow

Zhao et al
TVCG (VAST) '14
Application Area

- Investigative & Intelligence Analysis
  - Gather information from various sources then analyze and reason about what you find and know
  - Analyze situations, understand the particulars, anticipate what may happen

Problem Addressed

Help “investigators” explore, analyze and understand large document collections

- Articles & reports
- Spreadsheets
- XML documents
- Blogs
Visualization for Investigative Analysis across Document Collections

Law enforcement & intelligence community
Fraud (finance, accounting, banking)
Academic research
Journalism & reporting
Consumer research

"Putting the pieces together"

The Jigsaw Team

Carsten Görg
Zhicheng Liu
Youn-ah Kang
Jaeyeon Kihm
Jaegul Choo
Chad Stolper
Anand Sainath

and many others
Our Focus

- Entities within the documents
  - Person, place, organization, phone number, date, license plate, etc.
- Thesis: A story/narrative/plot/threat within the documents will involve a set of entities in coordination
Entity Identification

- Must identify and extract entities from plain text documents
  - Crucial for our work
- Not our main research focus – We use tools from others

Sample Document

Report: 20040610-4_10
May 14 2004

VANCOUVER, British Columbia — A Canadian immigration panel is considering whether accused environmental activist Tony Arrow can apply for refugee status in Canada.

Arrow, 30, who is wanted for firebombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether Arrow is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether Arrow is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.
Sample Document 2

Title: Proving Columbus was Wrong
Abstract: In this work, we show the world is really flat. To do this, we build a bunch of ships. Then we...
PI: Amerigo Vespucci
Co-PI: Vasco de Gama, Ponce de Leon
Organization: Northwest Central Univ.
Amount: 123,456
Program Mgr: Ephraim Glinert
Division: IIS
ProgramElementCode: 2860
Entities Already Identified

Title: Proving Columbus was Wrong
Abstract: In this work, we show the world is really flat. To do this, we build a bunch of ships. Then we...

PI: Amerigo Vespucci
Co-PI: Vasco de Gama, Ponce de Leon
Organization: Northwest Central Univ.
Amount: 123,456
Program Mgr: Ephraim Glinert
Division: IIS
ProgramElementCode: 2860

Connections

• Entities relate/connect to each other to make a larger “story”
• Connection definition:
  – Two entities are connected if they appear in a document together
  – The more documents they appear in together, the stronger the connection
Jigsaw

“Putting the pieces together”

- Computational analysis of document text
  - Entity identification, document similarity, clustering, summarization, sentiment
- Multiple visualizations (views) of documents, analysis results, entities and their connections
- Views are highly interactive and coordinated
Demo

- Car reviews
  - Text: Consumer’s comments
  - Entities: Various ratings (1-10), car features, other makes & models

Computational Analyses

- Document summarization
- Document similarity
- Document clustering by content
  - Text or entities
- Sentiment analysis
Application Domains

- Intelligence & law enforcement
  - Police cases
  - Won 2007 VAST Contest
  - Stasko et al, *Information Visualization* ’08
- Academic papers, PubMed
  - All InfoVis & VAST papers
  - CHI papers
  - Görg et al, KES ’10
- Investigative reporting
- Fraud
  - Finance, accounting, banking
- Grants
  - NSF CISE awards from 2000

- Topics on the web (medical condition)
  - Autism
- Consumer reviews
  - Amazon product reviews, edmunds.com, tripadvisor.com
  - Görg et al, HCIR ’10
- Business Intelligence
  - Patents, press releases, corporate agreements, ...
- Emails
  - White House logs
- Software
  - Source code repositories
  - Ruan et al, SoftVis ‘10
Potential Jigsaw Future Work

- Collaborative capabilities
- Improved evidence marshalling
- Present/browse investigation history
- Scalability upward
- Web document ingest
- Implement network algorithms
- DB import
- Wikipedia & Intellipedia
- Geospatial view
- Better timeline capabilities
- Reliability/uncertainty
- Other types of data
- Active crawling/RSS ingest
- Try it on display wall
- Deployment to real clients

Related Area of Interest

- Sensemaking
- A general term that has been used in a number of different contexts
  - E.g., How large corporations make decisions

- To me, ultimately about people working with data and information to understand it better
Sensemaking

Nice definition:

“A motivated, continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively.”

— Klein, Moon and Hoffman

*IEEE Intelligent Systems* ‘06

Alternate Definition

“The process of creating situation awareness in situations of uncertainty”

— D. Leedom, ‘01 SM Symp. Report

Situation awareness:

“It’s knowing what’s going on so you know what to do”

— B. McGuinness, quoting an Air Force pilot
This Topic

- I work on it a lot now

- Interested in getting more work in this area started

Project

- Presentation scheduling

- Any questions?
Upcoming

- Evaluation
  - Reading
    - (Will talk about Tableau too)

- Thanksgiving (no class)

Additional Material
Visual Analytics Partnership Disciplines

- Statistics, data representation and statistical graphics
- Geospatial and Temporal Sciences
- Applied Mathematics
- Knowledge representation, management and discovery
  - Ontology, semantics, NLP, extraction, synthesis, …
- Cognitive and Perceptual Sciences
- Communications: Capture, Illustrate and present a message
- Decision sciences
- Information and Scientific Visualization

And far more than homeland security

Multiple Techniques Contribute to Threat Assessment

Prediction
Cognition
Organization
Integration
Aggregation

Data
Information
Knowledge

Synthesis
Analysis
Content Management
Connect the Dots

Visual Analytics
Graph Matching
Pattern Analysis
Evidence Extraction
Link Discovery
Extraction
Uses Today

- Scientific Research
- Regulatory and Legal Communities
- Intelligence Analysis
- DOE and DOD
- Market Assessments
- Capability Analysis - Resumes
- Medical and Pharmaceutical Communities
- National Security and Law Enforcement
- Information Assurance, Web Analytics
- Technology Scanning, Asset and Intellectual Property Management

Capabilities Desired

- **Reduce the threat of terrorism** through the invention, development, evaluation, and deployment of technology to analyze masses of data in different formats and types, from different sources, with highly varying degrees of confidence levels, within time frames required for rapid decision making.

- Better understand the risks and vulnerabilities of our critical infrastructures, trade, ports, and immigration by combining sensor, computational and visual analytics technologies for in-the-field and strategic decision making.

- Enable **rapid visual communication technology for response teams** for clear understanding of the situation assessment and alternate options for response with geospatial, and multi-jurisdictional situations for WME and natural disasters.

- Ensure **effective information communication methods** and technologies throughout DHS missions of analysis, risk, levels of alerts, and response, in unwrappable levels of assessment with evidence and communication styles aimed within audience-centric applications for rapid understanding and action.

- Provide an **enduring talent base** of educated professionals supporting future developments requiring visual communication of integrated information and operational support missions.
Projects

- Let’s look at some of the research projects in this area

IN-SPIRE™ Visual Document Analysis

A “Thinking Aid” for advanced investigation of unstructured text

- Uncovers Common Topics in Large Document Collections
- Engaging Displays for Exploration
- Multiple Query and Search Tools
- Supports Real-Time Streaming Data
- Compatible with Foreign Languages
- Shows Trends over Time

http://in-spire.pnl.gov

Video
D-DUPE

System for entity resolution in large networks such as bibliographic collections

System does computational analysis and provides suggestions and user can augment and correct

WireVis

- Another VA investigative analysis project
- Helping Bank of America examine wire transfers of money
- Want to detect fraud and illegal actions

Thanks to R. Chang for some slide content
Particulars

- Who – Bank analysts
- Problem – Detect money laundering and fraud in wire transfers of money
- Data – Electronic records of wire transactions and information associated with each

Background

- Wire transfers of money can be complex
  - Have a “from” and “to” but often many “middlemen”
  - May not know who intermediaries are
- Millions of transfers per day occur
  - Vast majority are legal
- Bank has legal responsibility to report suspicious activities
Data

- Each transaction:
  - Money amount
  - Payer (could be third party)
  - Payee (could be an agent)
  - Potential intermediaries
  - Addresses of payer and payee, instructions, additional comments are optional

Challenges

No Standard Form...
When a wire leaves Bank of America in Charlotte...
The recipient can appear as if receiving at London, Indonesia or Singapore
Vice versa, if receiving from Indonesia to Charlotte
The sender can appear as if originating from London, Singapore, or Indonesia
Challenges

- Scale: BoA may do 200k transfers per day
- No international standard: loosely structured data
- Bad guys are smart and one step ahead
  - Detection tools are always reactive

System Overview

- Heatmap View (Accounts to Keywords Relationship)
- Search by Example (Find Similar Accounts)
- Strings and Beads (Relationships over Time)
- Keyword Network (Keyword Relationships)
Coast Guard Search & Rescue

Shows stations, incidents, response times

Visualize historical data and support "what if" explorations

Calculate risk assessments and then communicate visually

Malik et al
VAST '11

Other Examples
Many Others

- A number of nice examples shown earlier on Graph & Network visualization day
  - Wong: Graph Signatures
  - Perer: Social Action
  - etc.

Definitions

- Thinking\(^1\) - or reasoning - involves objectively connecting present beliefs with evidence in order to believe something else
- Critical Thinking\(^1\) is a deliberate meta-cognitive(thinking about thinking) thinking act whereby a person reflects on the quality of the reasoning process simultaneously while reasoning to a conclusion.
- Intelligence\(^1\) is a specialized form of knowledge, an activity, and an organization. As knowledge, intelligence informs leaders, uniquely aiding their judgment and decision-making. ...
“…the quality of our life and that of what we produce, make, or build depends precisely on the quality of our thoughts.”

**Elements of thought:**

- Points of View
- Purpose of the Thinking
- Implications & Consequences
- Question at Issue
- Assumptions
- Information
- Concepts
- Interpretation And Inference

* Foundations of Critical Thinking  www.criticalthinking.org

---

**Example: Heuer’s Central Ideas**

- “Tools and techniques that gear the analyst's mind to apply higher levels of critical thinking can substantially improve analysis... structuring information, challenging assumptions, and exploring alternative interpretations.”
Intelligence Process

Figure 2.1. Notional model of sensemaking loop for intelligence analysis derived from CTA.

Pirolli & Card
Intl Conf Intelligence Analysis '05

Wheaton
In preparation
Pain Points

- Cost structure of scanning and selecting items for further attention
- Analysts’ span of attention for evidence and hypotheses

Console
Document Cluster View

Clustered by document text or by entities
Summarized by three words

Document Grid View

User controls order and color
Sentiment analysis shown here
Calendar View

Showing connections between entities and dates

Circular Graph View

Connections between entities
Scatterplot View

Documents containing pairs of entities

Demo 2

- InfoVis & VAST papers
  - Text: paper title and abstract
  - Entities: author, keyword, year, conference, “concept”
Document Import

Various document formats with entity identification

Input Data Formats

- Text, csv, pdf, Word, html, Excel
- Jigsaw data file format
  - Our own xml

- DB?
  - Go to Excel
  - Go to text, transform to Jigsaw data file
Entity Aliasing
Room to Improve

- What Jigsaw doesn’t do so well now
  - The end-part of the Pirolli-Card model
    - Helping the analyst take notes, organize evidence, generate hypotheses, etc.
      (The Tablet is a first step)
    - Sometimes called “evidence marshalling”
  - Others have focused more on that aspect...
i2’s Analyst Notebook

Analyst's Notebook

- Leading commercial tool in this space (law enforcement and intelligence agencies)
- Large zooming workspace where analyst creates networks of entities and notes
- Often used to produce presentation or story of analysis done
Oculus’ Sandbox

- Flexible space for inserting text and graphics
- Objects can be dragged-and-dropped from their other analysis tools
- Flexible level of detail
- Flexible gestures for making space, inserting, etc.
- Assertions with evidence gates
- Reasoning templates
PARC’s Entity Workspace

Entity Workspace

- Tools for rapid ingest of entities from documents
- Can snap together entities into groups
- Can indicate level of interest in objects
- Four main view panels, with zooming UI
VT’s Analyst’s Workspace

• Uses spatial affordances from a large display area for benefit in sensemaking
• Analysts move around and arrange items (documents, entities, search results) to externalize the thinking process
  – Like working with pieces of paper on a conference table, but with computational capabilities

Andrews & North
VAST ’12