Problem-Driven Design Studies -- Money Donation to Public School

By Huaying Tian & Arthur Sun

Context
- In Brazil, elementary and secondary public education generally has poor quality.
- Every parent that can afford a private school does it, thus we have a huge number of private schools competing for students.
- Schools in São Paulo/Brazil have a huge number of private schools competing for students.
- Are students migrating from public to private schools?

Stakeholders
- Stakeholders: Students, Government, Schools

OUTLINE
- What are we going to do?
- Why do we need visualization?
- What data do we abstract?
- How to visualize the data?
- Components of our analysis and function

Problem-Driven Design Studies
- donation counted by grade
- Understand the data better
- What a great mass!
- data repository for our project.
- and also hosts the webpages and javascript libraries.
- charts and graphs based on the data.

What are we going to do?
1. Analyze data from a US-based non-profit organization website that allows individuals to donate money directly to public school.
2. Get the dataset and take a 9000-row table subset of the original dataset for our analysis purposes.
3. Create an informative analysis on the basis of the data attributes.
4. Visualize the data in an efficient and expressive way.

What data do we abstract?
- number of donations per year
- funding status
- total donations
- date posted
- resource type
- grade level
- school state
- date posted range
- resource types
- donation counted by poverty level
- funding status
- pie chart

How to visualize the data?
- pie chart
-Horizontal bar chart
- drop-down menu
-line chart

About the Data
- Educational Expenditure (public available, per year)
  - School name
  - School type (private/public)
  - School location (Latitude, Longitude, Postal Code, City, District)
  - Grade Level
  - Student Code
  - Student Grade
- Data size (per census year, we need at least two)
  - 5384431 Students
  - 28529 schools
  - 460 MB

Students Migration
Elementary and Secondary Schools in São Paulo/Brazil
Carolina Roman Amigo & Wenqiang (Dylan) Dong
CPSC 547 – Information Visualization
October 2015

T1 - Tasks for Schools
- Help schools identify migration pattern of students.
  - Are they losing more students than gaining?
  - To which schools are they going?
  - Is there any particular grade in which migration is more intense?
  - How their students migration compares to the other schools?

T2 - Tasks for Government
- Are there any areas of the state receiving more students than others?
- Are students migrating from public to private schools?
T1 - Help schools identify migration pattern of students

- map: school types
- categorical color map
- # of incoming & outgoing students
- line thickness
- interaction: select
- facet: juxtapose

T2 - Which areas of the state are receiving more students?

- map: (incoming-outgoing)%
  - diverging color map

Thank you!

Carolina Roman Amigo
carolamigo@gmail.com
Wenqiang (Dylan) Dong
wdong@cs.ubc.ca

VISUALIZATION OF YOUTUBE COMMENTS

- Doesn't support easy finding of entertaining comments.
- Emotionally draining arguments and trolls.

Task 1: Explore for entertaining comments.
- Doesn't support easy finding of entertaining comments.
  - Task 1: Explore for entertaining comments.
- Emotionally draining arguments and trolls.
  - Task 2: Identify arguments.
- Trolls = A single user (with little or no likes) being bombarded by multiple users
  - Task 3: Identify trolls.

Idea: A Bird's Eye View of the Youtube Comment Section

Neuron electrophysiology data visualization (Neuroelectro)

Presented by: Dmitry, Emily and Mike

Introduction

How does your brain work? It's complicated

Neuron electrophysiology is the study of the electrical properties of biological cells and tissues. In neuroscience, it includes measurements of the electrical activity of neurons, and particularly action potential activity.

What is our data?

Electrophysiology is the study of the electrical properties of biological cells and tissues. It involves measuring the electrical activity of neurons, particularly action potential activity.
What is our data?

How many neuron types are there? The debate has been ongoing for decades. We use enhanced NeuroLex.org definitions (~100 neuron types).

What is our data?

To summarize, we have per article:
1) Electrophysiology properties
2) Neuron types
3) Experimental metadata

We extract all of the above from published articles through text-mining and curation.

Current State

Mike

Problem Characterization

- We met with our stakeholder to ascertain high-level questions:
  - What do cells in different parts of the brain do?
  - How do experimental conditions affect electrophysiological measurements?
  - etc.

- We refined these into a few abstract tasks...

Task Analysis

- Discover relationships:
  - Neuron types (categorical)
  - Electrophysiological properties (quantitative)
  - Experimental conditions (quantitative and categorical)
- Narrow scope of analysis:
  - Select experimental conditions and electrophysiological properties to include
  - Filter by neuron type, electrophysiological property, and experimental conditions
We used an equal learning rate for all layers, which we adjusted manually throughout training. We initialized the neuron biases in the second, fourth, and fifth convolutional layers, with magnitudes proportional to the corresponding eigenvalues times a random variable drawn from a normal distribution with variance 0.01. We trained our models using stochastic gradient descent with momentum.

4.2 Dropout

Combining the predictions of many different models is a very successful way to reduce test errors. So every time an input is presented, the neural network samples a different architecture, “dropped out” in this way do not contribute to the forward pass and do not participate in back-propagation. This scheme approximately captures an important property of natural images, that both parts of the image, e.g. eyes and noses of dogs (layer 4, row 1, cols 1). Best viewed in electronic form. The compression artifacts are a consequence of the 30Mb submission.

Modern models learn features from data may have... 100s 1000s Millions of parameters

To understand features in a model, you used to just look at the fitted parameters. This makes it very difficult to understand what’s going on in your model!

My domain: Behavioural Game Theory understanding human behaviour in strategic situations

Feature discovery “by analogy”

Height: Tall

“Tall blondes”?

Hair colour: Blonde

Hand-crafted

Idea: derive distance between the output of learnt features and hand-crafted features.

In vision you can plot parameters directly.

... but this only works because of the visual structure of their models.
• I have data.
• I have hand-crafted features.
• I have a model.

Goal: analyze cameraman’s view

CameramanVis
Where camera should look?
Jianhui (Jimmy) Chen
CPSC 547 Pitch

Not always the same angle
-30°, 30° are almost opposite.

CameramanVis
Cameras for soccer games

Data

48-min (172, 800 frames) soccer game
Player positions on the playing ground
Camera angles of a PTZ camera
Both of them are noisy

Tasks

• Overview: camera angle, player position in playing ground
• Query1: given angle get player position distribution
• Query2: given player position get angle distribution
• Outlier detection: cameraman look at un-normal angle

Future Tasks for NLP Researchers

• Study link chains to associate link patterns with conversation types
  – Agreements
  – Debates
  – Off-topic/flamewars
  – Can we profile trolling?

Linking Sentences in Online Conversations

Jordon Johnson
M.Sc. Student
UBC Computer Science
jordon@cs.ubc.ca
Works and plays well with others

ConVis @ UBC

Visualizing Links Between Sentences
“Sentence B refers to sentence A”

GOAL: visualize/edit links
Link Attributes:
- Linked sentences
- Agreement value
- Sentiment value

Links may form chains that give additional insight into conversation structure

Data and Current Tasks

• Online conversations, annotated by humans to establish a “gold standard”
  – Not that good...
  – Use vis to make corrections and improve the gold standard
• (In progress) annotations by candidate linking algorithms
  – Use vis to evaluate the output

Future Tasks for NLP Researchers

• Study link chains to associate link patterns with conversation types
  – Agreements
  – Debates
  – Off-topic/flamewars
  – Can we profile trolling?
Thank you!

vidviz: scaling up video annotation

Julieta Martinez

The Internet

"the dark matter of the Internet" — Fei-Fei Li

"Plumeria Frangipani"
Interactive graphs using R Shiny
And maybe d3 if I have time.
Example: Gapminder world

- Goal: Create different types of visualizations for exploration of data.
- Fully interactive — linked highlighting, multi-faceted, animations!
- Learn what works, what doesn’t.
- Dataset: Gapminder. Wealth of information to explore, fairly easy to use.

Ken Mansfield, CPSC 547, Oct 22, 2015

R Shiny Examples
- Animation
- Colours
- Different types of Encoding

Interactivity

Expand on Gapminder

InfoVis Project Pitch
Kimberly Dextras-Romagnino

Present information to the masses in a visually appealing way to both inform and entertain

Storytelling: The Next Step for Visualization
- Robert Kosara
  - Presentation and Communication vs Exploration and Analysis
  - Data Visualization: Medium for Storytelling

What?
Why?
How?
How Does Your Body Manage Its Army?
By: Louie Dinh

Darwin until now: understanding of biology has been very qualitative

So….
Let’s Quantify! (With Sequencing)

Interesting problem: How does your body manage its army (the immune system)

When your body starts a civil war (autoimmune disease), how does that look compared to a normal person?

Sequencing data on blood cells is allowing us to visualize the your personal army

Very high dimensional. Highly correlated. How do we understand it?

If you “know” biology and find this problem interesting, come talk to me!
Containment Caching
- Reuse solutions to already solved problems
- Larger problems are most useful

Visualizing the Cache
- Several million cache entries - need to summarize!
- How much of the multi-dimensional solution space does the cache "cover"?
- How are the stations distributed across cache entries?
- What is the key usage distribution for a given auction trace? Is this similar between traces?
- How "distant" are individual cache entries from each other?
- Potential complication: Data is not public (I've signed an NDA)

Search Trends Visualization
- Multiple Related Keywords
- Example
  - HTML5
  - → jQuery
  - → HTML5 jQuery
  - → HTML5 Canvas
  - → SVG
  - → HTML5 SVG
  - → HTML5 SVG Canvas
  - → HTML5 jQuery Canvas

What & Why
- Data: Search Engine Statistics
  - Google, AdWords, Trends, Suggests
- Task: Given certain keywords, find related keywords that are:
  - Being searched more: Higher search volume
  - Getting more searched: Trending
- Rationale
  - Search Engine Optimization (SEO)
  - Search Keyword Efficiency

Example
- HTML5: 368000
  - → jQuery: 823000
  - → HTML5 jQuery: 720
  - → HTML5 Canvas: 22200
  - → SVG: 11000
  - → HTML5 SVG: 2400
  - → HTML5 SVG Canvas: 140
  - → HTML5 jQuery Canvas: 90

Example
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas
- [x, y] HTML5 HTML5 Canvas

Visualizing a SAP Network
Visualizing Movie Data

Yujie Yang, Ye Chen

I'm looking for a partner, Wanna join me?
© BCTV Malaysia

# Our heaven!
# A university to learn:
- what work(ed)/(s)
- what didn't/doesn't
Guess what?
I'm planning to visualize the heaven!

Movie Databases

- From movielens.org
- Ratings until Aug 2015!
- 2 versions are available:
  - Light
    - 700 Users
    - 9K Movies
    - 100K Ratings
  - Full
    - 230K Users
    - 30K Movies
    - 21M Ratings

Data set

Data set

# Features
+ Titles
+ Genre
+ Rating
+ Tags (Themes)
- IMDB ID!
+ IMDB Rating
+ Metascore
+ Year
+ Director
+ Writer
+ Awards
++ Many more…

Motivation

Used to be a TV presenter
Wrote/Directed ~200 episodes
Viewers' ratings mattered ($$)
• Keeping up w/ trends
• Touching popular genres
& 100s of more analyses

Challenges

# The DS needs some work
- Variable Recoding
- Data Cleaning
- Remote Data Fetching
# It's a programming project
- HTML/CSS/JavaScript
- JQuery/D3.js
- Java/JSP/JSF

Where?

Visualizing Uncertainty in Incomplete Election Data
CPSC 547 Project Pitch
Yasha Pushak

Visualizing Movie Data

Where?

What?
**Why?**

Explore the distribution of data.

**How?**

Help people to choose a movie.

---

**Dataset**

- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms...

**Dataset — Another Angle**

- 2M papers, 8M citations
  - title, authors, affiliations, year, venue, abstract
- 1M authors, 4M co-authorships
  - name, affiliations, #papers, #citations, H-index, key terms...

**Goal?**

- Make sense of whole dataset
- Explore a paper, a topic, an author, a venue...

---

**Law is Stuck**

Legal research is constrained by:
- High search costs
- Decentralized sources
- Institutional barriers
- Outdated Content Format
- Text-heavy, Static, and Linear
- Duplicate Search Costs
- Firm Work Product Silo
- High Client Costs
- Limited Access to Justice
- Content Barriers
- Private Content Paywall
- Solitary Search
- No Collective User Activity

**Knomos**

MAPPING A KNOWLEDGE NETWORK OF LAW

Visual navigation platform for big data research and collaboration in the legal industry

**Thank you**
Enhanced Legal Research Software

Team

- Adam La France (CEO)
- James Abney (CTO)
- Jesse Abney (COO)
- Craig McInnes (Systems Design)

Core Team

- History of industry leading software design applied to a unique visual platform for legal data research & collaboration

Key Partners

- Choice of D3 Vis Idioms
- Established Full Tech Stack w/ Room to Extend
- User Driven Design
- Networked Data: BCLaws & CanLII APIs

Why Knomos?

Big Bang

GoT Reference

Legal Research Usability Problems