Ryan McAdam
Project Proposal
Info 290: Social Computing

Overview

For the Social Computing final project I have elected to take a two-pronged approach with a survey paper. The primary subject of my paper will include a review of secondary source material – including relevant books, blogs, journals and conference papers. Additionally, I will explore the option of interviewing an expert in the field of social data and relate that discussion back to topics covered in the review of literature. What follows is a description of the topic that will motivate my research and a summary of the specific items that I anticipate will inform my discovery.

Background & Motivation

Before I cover the research details, I will briefly describe my background as it pertains to my interest in social computing. In my career as an interactive designer I have worked on projects which have included social data components. My responsibility was to visualize that data, often in a manner that supported or enhanced the narrative of the project. Increasingly, many designers are finding themselves in the role of “data visualizer”. Over the past decade there has been growing interest in this specialization within the field of design. And with the advent of more and more data (most notably from social networks) the role in presenting this data has become more relevant to unlocking understanding.

Another motivation for the survey paper will be to connect the lessons in Social Computing with other subjects in the I School. For example, I am currently enrolled in Quantitative Research Methods. Here, I see the potential to align knowledge of social computing, quantitative research and data visualization. I hope to more clearly define this potential through the course of the semester and to share those discoveries.

Significance to Social Computing

I have had direct experience with numerous questions that arise when deciding how to visually represent data:

- What is the appropriate fidelity for representing data? Can providing less information make the message clearer? Alternately, does too much accuracy cloud the intended message?
- What information makes sense to elevate in the visual hierarchy of a given design?
- How much of the graphic representation is intended to entertain and how much of it results in so-call chart junk?
- In the business domain, what obligation does a designer have toward promoting accuracy with regards to client objectives? What if a design is distorted but not the numbers?

I believe that a designer can better answer these questions if they are making informed decisions. And I believe those decisions can be informed with a fundamental understanding of the techniques and tools used to mine data. In other words, designers can benefit from expanding their perspective, and data visualization is a field that can merge the best of design with the best of data mining.

Problem Definition

But if data visualization suffers from specific blind spots, it can also be argued that pure data mining has it's drawbacks as well. From a designer's perspective, the "rawness" of the data is what prevents the information within from being effectively communicated. And if the story isn't accessible then data hasn't made the critical leap into being understandable.

Perhaps this tension is best summarized by Enrico Bertini in his blog, Fell in Love with Data:

> *. . . from the one hand data miners see visualization as a too soft discipline, lacking of enough formalism and with the big original sin of having very poor evaluation methods in its toolbox. From the other hand visualizers think data mining is too rigid and narrowly focussed on a plethora of insignificant small deltas to algorithms that nobody will ever understand.[1]*

Bertini goes on to make separate sets of recommendations to both groups. For visualizers, the tasks are to learn how to interpret and shape data, remain cognizant of human scale, and allowing the order that emerges from the data to take precedence over the design. For data miners, the tasks are to accept that by defining parameters for the model they are essential making design decisions. Miners can also use visualization to help peer into the "black box" of algorithms by incorporating it into the process of how they refine the model.

A central question of my survey will be: where are the opportunities for data visualization and data mining to share and benefit from their perspectives?

Approach

The topic of my survey will cover literature, discussion, debate and examples which illustrate the relationship between the methods by which social data is collected and the means by which that data is visually communicated. I will explore instances where data mining and data visualization are both complimentary yet competitive: complimentary in their mission to collect and reveal patterns within data; competitive in their approach to communicating what is discovered from the data. The transition between *knowledge discovery* and *knowledge visualization* will be a key area of interest.

Source Material Under Consideration

I have made a preliminary review of materials I believe may inform my survey. The following list is simply a starting point and will undoubtedly be updated as more sources surface. I would welcome any input on further readings.

Books

- *Information Visualization, Perception for Design*, Colin Ware
- *Envisioning Information*, Edward R. Tufte
- *Visualizing Data: Exploring & Explaining Data with the Processing Environment*, Ben Fry

Blog & Online Articles:

- http://infosthetics.com/
- http://www.gapminder.org/
- www.perceptualedge.com/
- http://ffctn.com/doc/ffunction-data_visualization-2010.pdf

Conference literature

- A review of pertinent literature from the International Conference on Information Visualization
- A review of pertinent literature from the Symposium on Information Visualization

Considered Interview(s)

Additionally, I intend to pursue an interview with a least one expert in the field of data visualization. Under consideration are:

- Enrico Bertini, a data visualization researcher at the University of Konstanz, Germany.
- Jer Thorp, a software artist, writer, and educator.
- Ben Fry, principal of Fathom, a design and software consultancy and one of the founders of Processing programming language.

Conclusion

I arrive at social computing from a visual communication background. Yet I am intensely curious about the tools available through data mining and how those tools can be applied to enhancing data visualization. Social data provides us with an unprecedented amount of material to discover. The task of visualization is to represent this data as meaningful information. Through my survey I hope to gain a better understanding of these challenges. I look forward to sharing the results.

[1] http://fellinlovewithdata.com/reflections/why-visualization-cannot-afford-ignoring-data-mining-and-vice-versa