Social and Cultural Computing

Fall 2013

The Graduate Center, CUNY (City University of New York)
365 5th Avenue, New York City.
Instructor: Lev Manovich

Course numbers: CSc 87100 / CRN 21789
Format: graduate seminar open to PhD and MA students.
Classes meet on Mondays, 11:45am - 1:45pm.
Room: 3306.

Course description:

The joint availability of massive cultural data sets (including social media and digitized cultural artifacts) make possible fundamentally new paradigms for the study of social and cultural activities and histories. While the recently emerged field of social computing started to explore some of these possibilities, we are only at the very beginning. In this seminar, the students will work together in groups to conceptualize and complete computational projects which use big social/cultural data.

Each team will identify new interesting and important questions which have not yet been asked, prepare data sets sets, analyze them, create visualizations, and professional looking project web site about the project. (The project can also result in an interactive app, or interactive artwork).

The students will have a choice of using any of the cultural data sets already assembled in the instructor's lab or gather new data sets from the web or other sources.

The only prerequisites for this class are 1) curiosity and interest in using large data and computational methods to ask questions relevant for society; 2) some background in one of the following area: statistics, multivariable data analysis, machine learning, data visualization, digital image processing, computer vision, text analytics, geospatial analytics, digital humanities.

Course plan:

September 9: class introduction
September 16: capturing data about the world and humans - technologies and examples of art projects:
September 23: social computing: selected papers and projects
September 30: cultural computing: selected papers and projects
October 7: data analysis concepts and techniques
October 15: presentation of student proposals for class projects
October 21: working on pilot project
October 28: working on pilot project / data analysis concepts and techniques
November 4: presentations and discussions of pilot projects
November 11: working on final project
November 18: working on final project
November 25: working on final project
December 2: presentation and discussion of final projects
December 11: fine tuning final projects

Examples of projects:

phototrails.net
livehoods.org

Grading:

Class participation: %20

Class presentations (papers in social computing or other relevant topics; project presentations): %20

Pilot project: %20

Final project: %40

CLASSES:

September 16:
capturing data about the world and humans - technologies and examples of art projects:

Digitization of cultural "artifacts" and recording of "digital traces" are just latest episodes in the longer history of recording the world and the humans as "data," and later as "media" (19th - 20th centuries).

Today we have four parallel processes: 1) digitally-based recording of the physical environment; 2) massive capture of human activities which are mediated by computers; 3) humans themselves sharing media and communication which is captured by social networks and accessible (in part) via APIs; 4) translation of older analog recordings into digital data.

New *data capture technologies* have major effects on society, economy, media, culture, and lead to new forms of art (when they become widely available):

19th-20th century:

photography, film, audio and video recording, radar, remote sensing technologies

Examples of art projects done with the recently developed new data captures technologies in the 21st century:

Lidar scanners (Wikipedia article)

Aaron Koblin, House of Cards video for Radiohead (interactive version). Non-interactives video version (YouTube). Making Of (YouTube). For other Aaron's projects, see http://www.aaronkoblin.com/work.html. Newest: Light Echoes.

GPS receivers

Masaki Fujihata, Field-Works@Alsace (2002)

milkproject.net (2004) (one of the earlist art projects which uses GPS)

Since location capture is now built into mobile phones and many popular social media apps, lots of projects explore this - for example: Eric Fisher, locals and tourists (set on

Flickr) (2010)

social media data

listening post (2002)
wefeelfine.org (2006)
Pulse of the Nation (2010)

motion capture

Ghostcatching, a digital art installation by Paul Kaiser and Shelley Eshkar uses motion capture of the dance performances by Bill T. Jones (1999).
Hacking and creative coding around Kinect platform.

High resolution digital photography

http://9-eyes.com/ (Jon Rafman)
http://www.wired.com/design/2013/08/22-beautiful-photographs-hidden-in-this-insane-150-gigapixel-image-of-tokyo/

HOMEWORK for sept 23

1. Explore free web and social media analytics tools

1) If you don't use Google Analytics on your site/blog, or have not heard of it, see this overview:

http://www.google.com/analytics/features/index.html

if you have an active web site, install Google Analytics and explore the results.

2) Try these Twitter analytics tools with your own account (if you have it), or accounts of other people:

http://twtrland.com/
http://www.twitonomy.com/

3) Other suggested tools to explore:

https://sumall.com/

2. Readings: Social Computing and Computational Social Science - Selected Papers

1) Life in the network: the coming age of computational social science. Science. 2009 February 6; 323(5915): 721–723.

2) Computational social science: Making the links. Nature, 22 August 2012.
Learn more about the research of people mentioned in the article by exploring their web sites and projects.

3) Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moo.
What is Twitter, a Social Network or a News Media? WWW 2010 conference.

4) Justin Cranshaw, Raz Schwartz, Jason I. Hong, Norman Sadeh.
The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. The 6th International AAAI Conference on Weblogs and Social Media. Dublin, Ireland, June 2012. See the project web site: http://livehoods.org/.

Check the press about the project to see how journalists saw it:
http://livehoods.org/press

5) Philip Salesses, Katja Schechtner, Cesar A. Hidalgo. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. Plus One, July 2013.

Project site:
http://pulse.media.mit.edu/

Coverage in Wired: http://www.wired.com/design/2013/09/can-quantifiable-emotions-change-the-design-of-cities/

HOMEWORK for sept 30

Readings: Cultural Computing - Selected Papers and Projects

1. Style Analysis (photos):

Naila Murray, Luca Marchesotti, Florent Perronnin. AVA: A Large-Scale Database for Aesthetic Visual Analysis. Computer Vision and Pattern Recognition (CVPR) - 2012

2) Jan C. van Gemert. Exploiting photographic style for category-level image classification by generalizing the spatial pyramid. ICMR '11 Proceedings of the 1st ACM International Conference on Multimedia Retrieval.

Background:

http://en.wikipedia.org/wiki/Vector_space_model
http://en.wikipedia.org/wiki/Bag_of_words
http://en.wikipedia.org/wiki/Visual_Word
http://en.wikipedia.org/wiki/Bag-of-words_model_in_computer_vision

2. Genre Analysis (music):

Anthony Liekens. Data mining musical profiles. 2007.

How We Understand Music Genres. http://blog.echonest.com/post/52385283599/how-we-understand-music-genres, June 7, 2013.

3. Exploring media collections (images)

Jing Yang, Jianping Fan, Daniel Hubball, Yuli Gao, Hangzai Luo and William Ribarsky. Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis. 2006 IEEE Symposium On Visual Analytics Science And Technology.

G.P. Nguyen, M. Worring. Interactive access to large image collections using similarity-based visualization. Journal of Visual Languages and Computing & Computing 19 (2008) 203–224.

4. **Patterns in social media activity:**

Mor Naaman, Hila Becker and Luis Gravano. Hip and Trendy: Characterizing Emerging Trends on Twitter. Journal of the American Society for Information Science and Technology (JASIST),

62(5), May 2011.

Mor Naaman, Amy Zhang, Samuel Brody, and Gilad Lotan. On the Study of Diurnal Urban Routines on Twitter. Proceedings ICWSM 2012.

HOMEWORK for oct 7

Projects to view:

Cinemetrics

http://prismbreakup.org/

Readings:

N.S.A. Gathers Data on Social Connections of U.S. Citizens. *New York Times*, September 28, 2013.

Artist Turns a Year's Worth of Tracking Data Into a Haunting Record. *Wired*, 07.22.13.

Review papers from topics (1) and (2) (previous homework) to prepare for their discussion.

HOMEWORK for oct 14

Prepare project proposal:

7-14 pages - Powerpoint or another format.

appr. 15 minutes presentation.

what? what do you want to do?

why? why this is relevant? Is this project going to advance our understanding of society and

culture? Are you using big data to do something which could not be done before? Are you using big data/computation to answer existing/old research questions, or ask new questions? Explain why your project is original.

previous work: can be papers or art projects or theories from any relevant social science/humanities field

how? What data and/or metadata will you use? if you can, specify key methods.

Recommended: take a look at the grant proposals for NEH Startup grants.

October 28:
Commonly used techniques for the analysis of social and cultural data sets

Distance measures and distance matrix

http://en.wikipedia.org/wiki/Distance_matrix

Section 2.3 in P*rinciples of Data Mining.pdf*

in R:
http://www.statmethods.net/advstats/mds.html

example:

```
d <- dist(mydata)
```

visualize distance matrix as heatmap:

```
mydata.scaled <- scale(mydata) # first scale the data
heatmap(as.matrix(dist(mydata.scaled)), Rowv=NA, Colv=NA)
```

application to text analysis

Sections 5.3.3 and 14.3 in P*rinciples of Data Mining.pdf*

MDS (Multi-dimensional scaling)

http://en.wikipedia.org/wiki/Multi-dimensional_scaling

Section 3.7 in P*rinciples of Data Mining.pdf*

in R:
http://www.statmethods.net/advstats/mds.html

example:

```
d <- dist(mydata) # first calculate distance matrix
mydata.mds <- cmdscale(d,eig=TRUE, k=2)
```

In Matlab:
http://www.mathworks.com/products/statistics/examples.html?file=%2Fproducts%2Fdemos%2Fshipping%2Fstats%2Fcmdscaledemo.html

PCA (Principal Component Analysis)

http://en.wikipedia.org/wiki/Principal_component_analysis

Section 3.6 in P*rinciples of Data Mining.pdf*

in R:

http://www.statmethods.net/advstats/factor.html

example:

```
mydata.pca <- princomp(mydata, cor=TRUE)
```

visualizing PCA results in R:
http://planspace.org/2013/02/03/pca-3d-visualization-and-clustering-in-r/

Examples of applications of these techniques for analysis of cultural data:

example of heatmap visualization of distance matrix:

See figure 2.1 -
James M. Hughesa, Daniel J. Grahamb, Daniel N. Rockmorea. Stylometrics of artwork: uses and limitations. SPIE 2010.

examples of PCA use:

Anthony Liekens. Data mining musical profiles. 2007. (Note: PCA in this article is calculated and plotted using R).

PCA for text analysis in Python:

https://www.clear.rice.edu/comp130/12spring/pca/pca_use.shtml

example of MDS use:

Jing Yang, Jianping Fan, Daniel Hubball, Yuli Gao, Hangzai Luo and William Ribarsky. Semantic Image Browser: Bridging Information Visualization with Automated Intelligent Image Analysis. 2006 IEEE Symposium On Visual Analytics Science And Technology

Selected books on practical data analysis using different languages:

Philipp K. Janert. Data Analysis with Open Source Tools. 2010.

Wes McKinney. Python for Data Analysis. 2012.

Tutorials - text analysis and visualization with Processing

HOMEWORK for dec 2 / dec 9

1) Read:

http://www.fastcolabs.com/3016160/what-hackers-should-know-about-machine-learning?partner=rss

2) Work on the text for your project.

It can be anywhere between 500 and 1000 words, and should address these 5 questions:

what? short one paragraph summary of the project.

why? why this is relevant? Is this project going to advance our understanding of society and culture? Are you using big data to do something which could not be done before? Are you using big data/computation to answer existing/old research questions, or ask new questions? Explain why your project is original.

previous / most relevant work: mention a few most relevant references - can be papers or art projects or theories from any relevant social science/humanities field.

how? Briefly explain how the project was done, including: what data you collected and how; algorithms / software used to analyze / visualize / map / sonify the data.

findings? Tell us the the interesting findings. Do they confirm our expectations and existing knowledge about the project's subject, or are they unexpected.

Please place your text in our dropbox folder, named as follows:

[your last name].project.description.docx

3) Prepare a working draft of your project.

It should be presented as a web page or a web site.

Use examples of projects in geo-spatial mapping we looked on Nov 25 for examples (I am putting a folder with these examples in the class Dropbox).

Include your name and email.

The project should include the text describing it (see above). If you don't want to include the complete text, you can include a shorter version - in this case, email me the complete text.

You will have time to refine your project after last class. I will need to receive your email with the link to the completed project before January 2, so I can enter course grades.