Crowdsourcing, Big Data, and Social Media in the Behavioral Sciences: Applications, Methods and Theory

Dec. 3 & 4, 2015

Institute for Mathematical Behavioral Sciences at UC Irvine

Conference Talk Schedule

Thursday, December 3, 2015

9:00am – 10:00am

Ulf-Dietrich Reips

University of Konstanz, Germany

(reips@uni-konstanz.de)

Research in and with Social Media

Abstract: In this presentation I will give an intro to behavioral and social science research with social media, and present examples from several types of social media (e.g. Flickr, Facebook, Twitter, Google search) and across several disciplines. I will then present some tools we have developed for this kind of research, along with example studies. Among other tools available from our iScience Server at http://iscience.eu I will present iScience Maps, a free Web service for researchers, available from http://tweetminer.eu/ that can be used to analyze tweets. Tweeting (sending messages via Twitter) and other microblogging has gained prominence as a way to broadcast personal messages. Location awareness and promptness provide researchers using the Internet with the opportunity to create "psychological landscapes" and timelines from the microblogged messages - that is, to detect differences and changes in voiced (twittered) emotions, cognitions, and behaviors. Another example is Social Lab, our "Open Source Facebook" available at http://sociallab.es that can be played with in learning about privacy at http://en.sociallab.es. The presentation will conclude with evolving and future scenarios of using social media in scientific work. Publications are available from http://tinyurl.com/reipspub.

Thursday, December 3, 2015

10:10am – 11:10am

Norbert Schwarz

Psychology, USC & David Hauser, Psychology, University of Michigan

(norbert.schwarz@usc.edu)

Online attention checks: Attentive Turkers and unintended consequences

Abstract: Because participant attentiveness is a concern for many researchers using Amazon's Mechanical Turk (MTurk) and related services, researchers frequently use online attention checks (aka instructional manipulation checks) to assess participants' attention to detail. We report on a series of experiments that (i) compares attention among subject pool participants and MTurk participants and (ii) identifies unintended effects of attention checks on participants' reasoning strategies. We find that MTurk participants (with a 95% approval rating for 100+ hits) pass attention checks at a rate of 90% and above, whereas subject pool participants show much lower attention levels, with pass rates between 30% and 40% online as well as in the laboratory, even when an experimenter is present. The better performance of MTurk participants was also observed on novel attention checks that could not be identified based on previous experience. We conclude that Turkers who meet standard reputation-based selection criteria are aware that their attention may be checked and do, indeed, pay more attention. Other experiments highlight, however, that attention checks come at a cost: they alert participants that the study may include "trick questions", which affects performance on tasks that benefit from suspicion. For example, participants performed better on tasks from the Cognitive Reflection Test when they were preceded rather than followed by an attention check. However, attention checks do not influence the emergence or size of context effects in questionnaires, including effects of question order or scale format.

Thursday, December 3, 2015

11:20am – 12:20pm

Patricia Greenfield

Psychology, University of California, Los Angeles

(Greenfield@psych.ucla.edu)

Cultural evolution in China and the U.S.: Using the Google Ngram Viewer to study implications of social and political change for cultural values and human development

Abstract: Longitudinal studies using "big data" supplied by the Google Ngram Viewer reveal changing cultural values that reflect social change. Shifting word frequencies in cultural products (millions of books) reflect, for example the rise of urban populations and the decline of rural populations in the United States and the U.K. (from 1800-2000) and China (from 1970 to 2008). In these periods, words indexing individualistic values adapted to urban, more formally educated, and wealthier environments rose in frequency, whereas words indexing collectivistic values adapted to rural, less formally educated, and poorer environments declined in frequency. Is change on the cultural level, assessed by means of massive data analysis of cultural products, reflected in change on the individual level? An interview study in China comparing child development and socialization across three generations answers this question affirmatively.

Thursday, December 3, 2015

1:30pm – 2:30pm

Mark Steyvers

Cognitive Sciences, University of California, Irvine

(mark.steyvers@uci.edu)

Combining Human Judgments

Abstract: In this research, we build on ideas from cognitive science and machine learning to build aggregation models that combine human judgments in general knowledge and prediction tasks. We propose that a successful approach to the aggregation of human judgment requires a cognitive modeling framework that explains how individuals produce their answers, and that also allows for individual differences in skill and expertise of participants.

Thursday, December 3, 2015

2:40pm – 3:40pm

Siddharth Suri

Microsoft, NY

(suri@microsoft.com)

Crowdwork's Invisible Engine: Valuing the Organic Collaboration that Drives Crowdsourcing Labor Markets

Abstract: This presentation shares key findings from joint research with Mary Gray that examines workers' experiences of crowdsourcing-for-hire labor markets or what we call crowdwork—paid, short contract tasks distributed to a large pool of online workers (aka the 'crowd') through open call platforms that might otherwise go to a full time employee. Combining 21 months of ethnographic fieldwork and interviewing with computational analysis of backend metadata, we compare the cases of workers in India and the United States, to understand the cultural meaning, political implications, and ethical demands of crowdwork. This talk lays out the evidence we have that, despite the designs of crowdsourcing to maximize efficiencies through atomized, autonomous workflows, the most active crowdworkers are not the independent workers they are assumed to be. Instead, workers collaborate extensively to address both technical and social needs generated by the platforms they work on. Specifically, crowdworkers collaborate with members of their networks to 1) manage the administrative overhead associated with crowdwork, 2) find lucrative tasks and reputable employers and 3) recreate the social connections and support often associated with brick-and-mortar work environments. Our work suggests that crowdsourcing's successes are driven as much by the invisible labor of organic collaboration among workers as the matching and routing capacities of an API. While our study includes examples of what we call "engineered collaboration," the persistence of organic collaboration even in these systems suggests that there's an untapped reservoir of productivity and better work conditions to be found in models of facilitated organic collaboration. We argue that before we can improve the technical capacities of crowdsourcing systems, we need methodologies that help us get a clearer sense of the people doing this work, what it means to them, and how it fits into their daily lives. In short, we should recognize that crowdsourcing systems are not, simply, technologies. In the case of crowdwork, crowdsourcing platforms are sites of employment with complicated social dynamics that, ultimately, hold value.

Thursday, December 3, 2015

3:50pm – 4:50pm

Niloufar Salehi, Lilly Irani, Michael Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier

Computer Science and The Human Computer Interaction Group, Stanford University

(niloufar@cs.stanford.edu)

We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers

Abstract: By lowering the costs of communication, the web promises to enable distributed collectives to act around shared issues. However, many collective action efforts never succeed: while the web's affordances make it easy to gather, these same decentralizing characteristics impede any focus towards action. In this paper, we study challenges to collective action efforts through the lens of online labor by engaging with Amazon Mechanical Turk workers. Through a year of ethnographic fieldwork, we sought to understand online workers' unique barriers to collective action. We then created Dynamo, a platform to support the Mechanical Turk community in forming publics around issues and then mobilizing. We found that collective action publics tread a precariously narrow path between the twin perils of stalling and friction, balancing with each step between losing momentum and flaring into acrimony. However, specially structured labor to maintain efforts' forward motion can help such publics take action.

Friday, December 4, 2015

9:00am – 10:00am

Kimberly A. Jameson[1], Sean Tauber[1], Prutha S. Deshpande[2], Stephanie M. Chang[3], and Sergio Gago[3]

[1]Institute for Mathematical Behavioral Sciences, [2]Cognitive Sciences, and [3]Calit2, University of California, Irvine

(kjameson@uci.edu)

Crowdsourcing the transcription of archival data

Abstract: We investigated using Internet-based procedures to convert information from a large handwritten archive of ethnographic survey data into a computer addressable database. Rather than manually transcribe the archive's estimated 23,000 pages of handwritten data, we sought to develop novel approaches to first collect crowdsourced transcriptions of the handwritten data, and then to use an innovative variation of Cultural Consensus Analysis (CCT) to objectively aggregate crowdsourced responses based on a formal process model of shared knowledge. Results of the investigations clarify several unanswered questions on this issue. Uncertainties the research addressed include: (1) whether crowdsourced tasks were practical as a method for automating the transcription of the archive's handwritten material, (2) whether responses from perceptually-based

tasks inherent to transcribing handwritten documents could be analyzed with CCT and, (3) if CCT analyses were extendable and appropriate as a model of the transcription challenge, was it then found that the output of CCT analyses produced accurate "answer-key" estimates that could serve as "correct" transcriptions of the archive's data. Our results address these issues and convey how CCT modeling and analyses can be modified and made appropriate for aggregating these data. Implications of these analyses and uses of CCT in large-scale crowdsourced data collection platforms are discussed.

Friday, December 4, 2015

10:10am – 11:10am

Michael D. Lee

Cognitive Sciences, University of California, Irvine

(mdlee@uci.edu)

Making sports predictions by applying cognitive models to crowd-sourced data

Abstract: We consider crowd-sourced data for a variety of sports predictions, including the rankings of NFL teams, the win-loss ratios of MLB, NBA, and NFL teams, the outcome of the FIFA World Cup, and results for the NCAA March Madness basketball tournament. The goal is to aggregate the predictions made by individuals, creating a "crowd" prediction that performs well. The key innovation is to rely on cognitive models to do the aggregation. This is motivated by the idea that the behavioral data being aggregated are the outcomes of cognitive processes, and the knowledge they express needs to be understood and measured in terms of those processes. In most of our applications, we are able to demonstrate useful improvements in crowd predictions coming from this cognitive modeling approach.

Student Poster Session and Lunch from 11:30am – 1:00pm

Friday, December 4, 2015

1:00pm – 2:00pm

Alexander Ihler

Computer Science, University of California, Irvine

(ihler@ics.uci.edu)

Computational Choices for Crowdsourcing

Abstract: Obtaining expert information has always been a major bottleneck for machine learning methods. "Crowdsourcing" techniques such as Amazon's Mechanical Turk have become a popular mechanism to access the power of human intelligence, for example to label large datasets. However, this raises the computational task of properly aggregating the crowdsourced predictions provided by a collection of unreliable and diverse annotators. We transform crowdsourcing into a standard inference problem on a graphical model, and apply powerful algorithms such as belief propagation (BP). We show that several existing methods are special cases corresponding to particular modeling choices, while with more careful choices, our simple BP method performs competitively with state-of-the-art methods on both simulated and real-world datasets. We use our framework to study the effect of modeling choices, inference algorithms, and the value of items whose true answers are known ("control" items) in crowdsourcing problems. We also give some theoretical results on the usefulness of control items under different scenarios, and provide a simple rule of thumb for practice. As a byproduct, we also provide theoretical analysis of the accuracy of different consensus methods.

Friday, December 4, 2015

2:10pm – 4:10pm

Ulf-Dietrich Reips

University of Konstanz, Germany

(reips@uni-konstanz.de)

Tutorial Session: Tools and methods in crowdsourcing and Internet-based experimenting

Abstract: In this presentation I will provide guidelines and overviews of techniques, methods, and tools for crowdsourcing and Internet-based experimenting that provide solutions to many of the methodological challenges in data collection via the Internet. The challenges discussed include issues in selection of and access to social media platforms, design, security, recruitment, sampling,

self-selection, multiple submissions, measurement scales, reactance-free question design, response time measurement, dropout, error estimation, data assessment, data handling, and data quality. For example, in creating items for a Web questionnaire it is often forgotten to add a neutral option ("please choose") in drop-down menus. Thus, one of the answer options is automatically selected, even if the respondent skips the item. Forcing a response has been shown to create psychological reactance and thus reduce data quality. Another pitfall - and opportunity - is the higher likelihood for attrition or dropout on the Web, compared to lab-based studies. Tools that automatically implement optimal solutions, e.g. the Web experiment generator WEXTOR (http://wextor.org) will be presented, along with other tools available from the iScience Server at http://iscience.eu. We will create an example Web experiment during the tutorial session, and take first steps in conducting a crowdsourcing study. Among other methods that have been developed in Web methodology and Internet science, I will explain the one-item-one-screen (OIOS) design, the seriousness check, sub-sampling procedures, and the multiple site entry and discuss empirical results from investigations into the validity of several of the techniques. Some time will be devoted to the "how to" in data collection in social media (e.g. Twitter) and to issues with privacy and self-disclosure. The presentation will conclude with evolving and future scenarios of the Internet's influence on scientific work.

Friday, December 4, 2015

4:20pm – 5:20pm

Gary H. McClelland

University of Colorado Boulder

gary.mcclelland@me.com

Visualization for Big Data and Internet Research

Abstract: Big Data and internet experiments usually generate more observations than can be usefully summarized in tables and quantitative summaries. Quality visualizations are required. I present some early success stories for Big Data from the late 19th Century, which Michael Friendly calls the Golden Age of statistical graphics. Then I examine how we've lost our way with the graphs that populate our contemporary journals and Powerpoint presentations. Finally, using some examples from internet experiments, I examine how we might again produce graphics worthy of our data and the effort it took to collect them.

Student Poster Session on Friday, December 4, 2015 from 11:30am – 1:00pm

Nathanial Benjamin[1], Sergio Gago[1], Ian Harris[2], and Kimberly Jameson[3]

[1]Calit2, [2]Computer Science, and [3]Institute for Mathematical Behavioral Science, University of California, Irvine

(nbenjami@uci.edu)

An Affordance Based Approach to Large Data-Set Navigation

Abstract: Relational databases are difficult to traverse without advanced knowledge and a fair amount of abstraction. Even for familiarized database users, the sheer volume of data can be overwhelming given the number of possible (and sometimes tenuous) linkages between objects. To address these difficulties we examine a large dataset, consisting of color categorizing data from a large number of participants responding in different languages (The Robert E. MacLaury Color Categorization Archive). Our aim is to make the archive's data accessible to researchers from a wide variety of disciplines, including students who may not have the technical knowledge or time required to learn to navigate the database effectively. To address this need, we aim to design and implement an application that provides an intuitive interface to better access the database's technical knowledge (i.e. Server Query Language). There are several interface tools that might be appropriate for this goal (buttons usually have an immediate action, radio buttons inform people to select only one option, whereas check-boxes imply the ability to make multiple selections). By tracking users' responses, an interface can use contextual cues to exclude invalid user actions. It is also possible to optimize exploration of vast amounts of information by using those same contextual cues to refine the displayed information. We have employed this approach to implement an intuitive database navigation that avoids the typical user learning curve found in many existing systems. Our results may be generalizable beyond the current project.

Stephanie M. Chang[1], Sergio Gago[1], Ian Harris[2], and Kimberly Jameson[3]

[1]Calit2, [2]Computer Science, and [3]Institute for Mathematical Behavioral Science, University of California, Irvine

(changsm1@uci.edu)

Designing Crowdsourcing Methods for the Transcription of Handwritten Documents
Abstract: Emerging approaches of online collaborative research permit novel methods that enlist the power of the Internet and human subject processing to rapidly convert handwritten forms of

archival information into data-addressable file formats. This project develops and tests crowdsourcing methods for transcribing a large corpus of natural categorization phenomena from a large number of ethnolinguistic cultures. We designed and implemented a flexible and scalable Web application to determine the feasibility of using crowdsourcing as a method for transcribing handwritten documents. The gathered information will ultimately populate an online archive of color categorization data from over 116 indigenous languages. Efficient strategies are required for transcribing, digitizing, and verifying the data obtained from our task designs. Vital features of the Web application include provisions for creating task modules appropriate for the documents needing transcription, permitting data gathering in efficient and accurate manners, and devising methods of optimally organizing, exporting, and storing the collected data for future use (e.g., integration into a relational database). The application might generally serve as a software tool for transcribing various types of handwritten data. With this form of crowdsourced data collection, we aim to develop an automatic, accurate, and efficient tool for transcribing documents for public dissemination and use by the research community.

Prutha S. Deshpande[1], Sean Tauber[2], Stephanie M. Chang[3], Sergio Gago[3], and Kimberly A. Jameson[2]

[1]Cognitive Sciences, [2]Institute for Mathematical Behavioral Sciences, and [3]Calit2, University of California, Irvine

(pdeshpan@uci.edu)

A Cultural Consensus Theory Analysis of Crowdsourced Transcription Data

Abstract: Cultural Consensus Theory (CCT) is a formal process model used to derive information from a shared knowledge domain when the "correct answers" are uncertain. We applied CCT to the analysis of internet-based crowdsourced responses, collected with the aim of transcribing a large handwritten cross-cultural survey dataset (MacLaury, 1997). Prior to this study, it was unknown whether CCT could serve as an appropriate model for assessing the accuracy of survey information collected via crowdsourced procedures, or if it could be useful for converting handwritten documents into data-addressable files. This study provides a preliminary investigation of such uses of CCT. We compare a standard Bayesian CCT model (Oravecz et al., 2014) with an alternative implementation of CCT that permits the use of multiple-choice/free-response data. The alternative model makes the non-standard assumption that response bias is based on perceptual confusability of stimuli and response options for each question, and therefore varies across items but not across individuals. Accurate transcription results were obtained with both implementations for dichotomous-type questions, supporting the continued use of the alternative implementation. Further analyses of multiple-choice/free-response data found all of the alternative model's criteria for fit to be satisfied, indicating the appropriateness of CCT for such data. Interestingly, two variations found accuracy of consensus solutions greatly improved by the inclusion of data from

expert respondents, and by considering data of only a subset of participants who responded to the task with high confidence. Implications of these analyses and uses of CCT in large-scale crowdsourced data collection platforms are discussed.

Yang Jiao[1], Sergio Gago[2], Ian Harris[1], and Kimberly Jameson[3]

[1]Computer Science, [2]Calit2, and [3]Institute for Mathematical Behavioral Science, University of California, Irvine

(yjiao7@uci.edu)

Optical Character Recognition of Handwritten Tabular Data

Abstract: Optical character recognition (OCR) technology transcribes images containing typewritten or handwritten script into machine-coded formats such as a text file or spreadsheet. One of the main advantages of OCR is to reduce the overall time and human effort used for transcribing text images into editable text content. OCR has already been applied in various domains including machine translation, text transcription, and text mining. OCR research papers have been published for recognizing Latin, Chinese and Japanese characters, but research on OCR handwritten transcription is not yet sufficient. For tabular data, image features such as text region segmentation, skew correction and line detection can be major factors that influence OCR recognition accuracy. We made use of Tesseract, an existing open source OCR engine, to analyze image representations of handwritten tabular data samples from the Robert E. MacLaury color survey. We developed new methodologies for segmentation and post-processing tabular data to enhance the performance of extracting and categorizing tabular data. Our design presents an approach that overcomes two unique complexities to tabular data. The first problem is the identification of a table's text cells (row and column) so that the results can be presented on spreadsheet corresponding to the original tabular data. The second problem is the visualization of confidence level of recognition results in color-coded format to visualize two-dimensional tables. Our research aims to develop a software tool that generates a spreadsheet containing the original tabular image with cells colored by the level of confidence for the classification of each cell.

William Leibzon

University of California, Irvine

(wleibzon@uci.edu)

Analysis of GitHub Social Network and Collaboration in Software Development

Abstract: This research looks at organizations of teams in software development area. GitHub, very popular collaborative software development website with social-network like features is used as a source for data collection. Using network analysis methods several open-source software development projects are evaluated and social network formed by collaboration and communication links between users with ties to each project are shown to have scale-free properties. A core development group for each project is re-constructed out of larger network and a network metric introduced that allows to measure how much collaboration core members of the project have with each other, which can be used to ascertain if a project is healthy and likely to be successful.

Yiran Wang[1], Gloria Mark[1], Melissa Niiya[2], Stephanie Reich[2], and Mark Warschauer[2]

[1]Department of Informatics, and [2]School of Education, University of California, Irvine

(yiranw2@uci.edu)

An Investigation of Stress and Mood Associated with Digital Media Use in Everyday College Life

Abstract: The relationship between how the millennial generation uses information and communication technologies (ICTs) and their well-being has been the subject of increasing studies. To empirically ground the discussion, we used precision tracking that includes automatic computer and phone logging, biosensors, and daily surveys to investigate college students' digital behavior in their in situ environments. We report on two studies in this poster: study 1 focuses on the extent to which college students multitask and how this behavior is associated with stress; study 2 seeks to provide a grounded understanding of how college students embed social media in their lives and the effect of social media use on their mood.