

Matching Data and Interpretation: Towards a Rosetta Stone Joining Behavioral and Survey Data

Drew Margolin¹ Yu-Ru Lin² Devon Brewer³ David Lazer^{1,2}

¹College of Social Sciences and Humanities, Northeastern University, Boston, MA 02115, USA

²College of Computer and Information Science, Northeastern University, Boston, MA 02115, USA

³Interdisciplinary Scientific Research, Seattle, WA, USA

Abstract

This position paper argues for the establishment of a "Rosetta Stone" that maps behavioral signatures of population behavior to meaningful social categories. We propose a method for accomplishing this through the deployment of behavioral tracking combined with survey questionnaires. The goal is to identify salient patterns of behavioral activity and ask subjects to "explain" these patterns through their survey responses. As a case study, we provide analyses regarding calling patterns related to emergencies, such as bombings and hurricanes.

Introduction

The continuing invention, improvement, and dissemination of communication technology has substantially improved the behavioral data available for research (Lazer et al. 2009). Increasingly, it is possible to take large, real-time streams of collective behavior and analyze them to produce a new understanding of life in a city or region. Yet the meaningful interpretation of these data for civic and other policy-making remains elusive. Algorithms can identify statistically regular patterns, but applying knowledge of these patterns to inform intelligent policy decisions or interventions remains difficult.

One of the reasons for this difficulty is that while these data are concrete indicators of important behaviors in civic life, such as the communication of information or physical migration, the social goals of civic institutions extend beyond producing or discouraging specific behaviors and include softer, more elusive projects that combine behavior with psychological and social components, such as engagement, satisfaction, or civic pride (Campbell and Kwak 2010). For several decades social scientists have attempted to measure these "softer" variables through direct engagement with citizens via surveys, interviews, and focus groups. These methods contain many weaknesses, however, including expense, responses biases, and limits on longitudinal design. Since behavioral data overcome many of these weaknesses, there is much hope for what can be accomplished using these new behavioral data streams.

We suggest that these sources of data are complementary rather than in competition. The limitations of each are nicely

Copyright © 2013, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

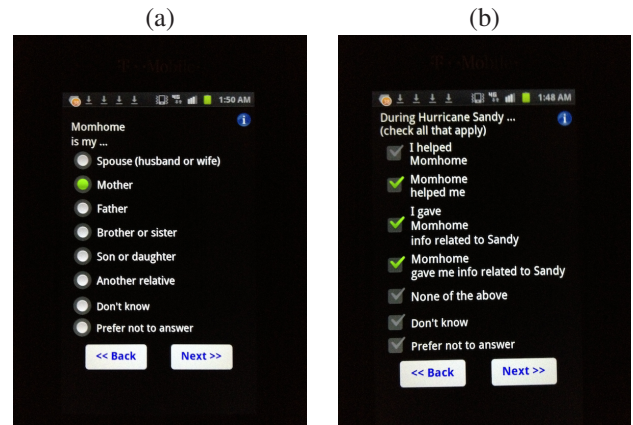


Figure 1: Screen shots from the Sandy Survey Android App.

balanced by the strengths of the other. The important research task is to develop a useful, practical methodology for combining the insights for these approaches.

In this paper we argue for one particular approach – a multi-method technique for building a "Rosetta Stone" that maps behavioral data to meaningful social theoretic categories that are relevant for public and policy concern. There are two key aspects to our approach. First, we examine behavioral and subjective responses to conditions or circumstances where it is reasonable to expect a large number of people are responding to a common circumstance. Such common circumstances include things such as weather emergencies, important breaking news, or other significant social events. This allows us to aggregate the responses under the common guiding rubric of this condition. Second, we extract behavioral signatures and design survey questions relevant to this condition with the explicit aim of mapping one to the other. We ask "what does [survey response x] look like in [behavioral data y]?"

To illustrate our approach, we provide an example analysis centered on "emergency situations," such as natural disasters or terrorist attacks. During emergency situations, normal routines are broken. Individuals activate close ties and relationships to access and provide help, comfort, and information (Shklovksi et al. 2010). These situations thus provide an opportunity to identifying the most important relationships in an individuals repertoire through behavioral data,

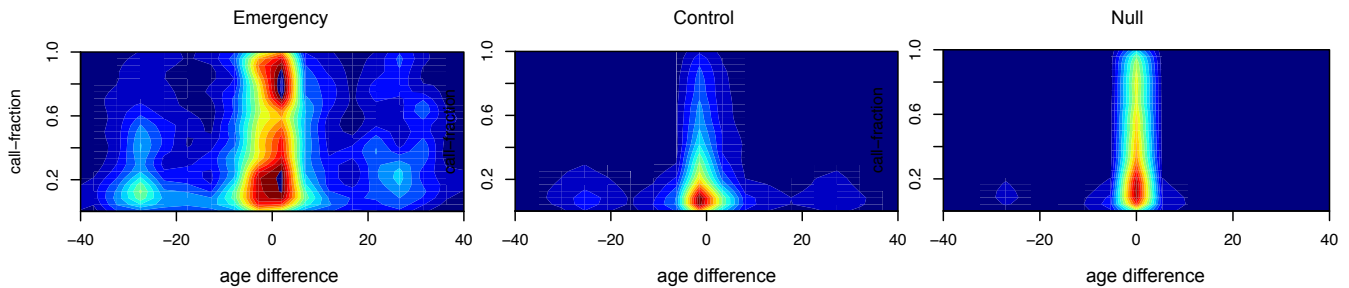


Figure 2: The intensity of call volume between individuals of different ages under normal times (Null) and during a planned event (Control). All periods show a propensity for individuals to call those close to their own age. During the emergency, however, activity increases for contacts who are 20-30 years older or young than the caller.

such as phone calls or mobility patterns. The salience of the event also permits the use of survey methods that ask people to describe or elaborate on their behavior.

Our illustrative case study draws on three sources of data from two “emergency situations.” The first emergency situation is an explosion in the downtown area of major European city. Call data records (CDRs) from this emergency were used to detect call patterns that deviated from “normal” patterns. The second emergency situation is the recent Hurricane Sandy that swept over the northeastern United States. Participants downloaded an Android app called the “Sandy Survey” which mined their call logs for the highest intensity contacts during the period around Hurricane Sandy. The app provides a survey in which individuals were asked to identify the role relationships (e.g. parent, sibling, friend) associated with each contact as well as the nature of the communication (e.g. to provide information, offer help).

Research Method

By combining the CDRs and the app survey we hope to be able to answer questions such as:

- What role relationships do people activate during disasters?
- What are the statistical signatures, within call log and CDR data, of such relationships?
- How do such activations relate to an individual’s success in obtaining help or information during an emergency?
- To what extent do other behavioral data, such as volume or diversity of calls, reveal these differences?

For example, CDRs from the period around a bombing in a major European city suggest that individuals that tend to be called most frequently during non-emergency periods are even more likely to be called during emergency periods. There also appears to be a tendency for specific relationship types to be activated. In particular, there is evidence suggesting that parents call children and vice versa as shown by the 20-30 year gap in age between typical callers (see Figure 2).

These behavioral patterns are suggestive of social patterns that make intuitive and theoretical sense: during emergencies people tend to reach out to loved ones and close friends to offer help or information. Parents and children tend to call

one another. Yet this apparent corroboration between behavior, memory, and intuition is speculative. With responses to the Sandy Survey, however, these inferences can be confirmed to a greater degree. Respondents explicitly identify their relationship to the people in their logs, allowing us to determine whether these “most called” individuals are in fact parents, children or other close ties of a particular kind.

Inferences can also flow in the other direction. In the Sandy Survey, individuals are asked whether they received sufficient help and information during the Hurricane. The call patterns for individuals who were satisfied with the help and information they received can then be statistically compared to individuals who were not satisfied. Significant differences in call patterns can then be applied to the CDR dataset for evidence of locations or individuals with particular characteristics that may have been underserved by authorities or their communities immediately following the bombing.

Discussion and Future Work

This paper outlines our approach to creating a “Rosetta Stone” map. Analysis from our illustrative case study demonstrates the potential insights from this approach. The goal is to extend this approach to other settings and conditions beyond the “emergency situations” described here. Examples of potential targets include sporting events, rock concerts and or media events.

Acknowledgements

This research was supported by NSF Award #1125095.

References

- Campbell, S., and Kwak, N. 2010. Mobile communication and civic life: Linking patterns of use to civic and political engagement. *Journal of Communication* 60:536–555.
- Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabasi, A.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. 2009. Social science. computational social science. *Science (New York, NY)* 323(5915):721.
- Shklovksi, I.; Burke, M.; Kiesler, S.; and Kraut, R. 2010. Technology adoption and use in the aftermath of hurricane katrina in new orleans. *American Behavioral Scientist* 53(8):1228–1246.