'Big Data' for Development

Daniel Björkegren, Stanford University, danbjork@stanford.edu
16 June 2010
CEPR/AMID Summer School

The recent spread of information technologies in developing countries has generated massive amounts of high frequency data. Analysis of this 'big data' can open up new perspectives of developing economies, perspectives that can lead us to better understand why poverty persists and what can be done to best help the poor.

One promising source of data is call detail records from mobile phone operators. These records can be used to infer individual behavior at an extremely fine level of detail, providing information on consumption of services over time, electronic transfers of credit, interactions within social networks, and mobility (based on cell tower data). For richer analysis this data can also be linked to phone surveys (e.g., Blumenstock, Gillick, and Eagle 2010), or other data sets, using phone numbers or geographic area as identifiers.

I am collaborating with Nathan Eagle (MIT/Santa Fe Institute) who has access to approximately 3 years of mobile phone records, representing 1.5m callers, from an operator in Rwanda, and am exploring the following research directions:

1. Credit scoring for unbanked borrowers

Although unbanked households lack the formal records needed for traditional credit scores, many have maintained a rich history of interaction with a formal institution over an extended period of time—their mobile phone activity, recorded by their operator. Even with prepaid plans, operator records can yield rich information about individual behavior and social networks.

If indicators derived from mobile phone data are predictive of creditworthiness, they could be assembled into credit scores that could help MFIs offer loans to unbanked households at lower rates, without the need for collateral. (For example, perhaps a good borrower keeps their phone topped up so they have credit in case of emergency, whereas one prone to default allows it to run out, depending on others to call them.)

Although a credit scoring method in itself is not interesting from an academic economic perspective, it may help bring an MFI onboard and provides a context for investigation of more interesting academic questions.

I am currently trying to set up a partnership with an MFI for this study, which I envision will have at least two parts:

I. **Policy:** assemble and test a credit score based on the most predictive metrics

II. **Economics:**

a) Intervention: Expansion of credit. As a follow up, the partner MFI could offer credit to a new set of unbanked households. Results could validate the credit scoring system and evaluate aspects of microfinance such as impacts on unbanked households and on social networks.

b) Borrowing in time. Development economists rarely have high frequency evidence of behavior in time. Merged mobile phone records and loan payment records could help answer questions such as:

• How does loan behavior interact with a borrower's consumption and mobility?

Posted at http://www.cepr.org/meets/wkcn/7/783/papers/Bjorkegren.pdf

- How do shocks interact with social networks, loan repayment, and mobile phone usage?

2. Impact Evaluation

Mobile phone data could be used to measure the impact of programs, or to investigate the effects of natural experiments. For example, the Rwanda Development Bank (BRD) has distributed several thousand handsets through a subsidy/credit scheme with the goal of increasing mobile phone penetration (Butera 2010). If there are exogenous discontinuities in distribution or how people were informed about distribution, the impact of the program could be evaluated as a natural experiment.

3. Social Network Analysis

Mobile phone records provide a promising platform for tracking learning across the social network. There are many different behaviors that are visible in call data and may be learned through a social network. For example, use of new modes of interaction (credit transfer service, mobile payments), SMS, and interaction with known third party phone numbers (SMS banking).

The data could be used to describe how individuals learn about technologies. Call data can be used to derive a predicted social network for each individual and track the spread of feature utilization through the network. While many studies have explored diffusion of a single technology over a network over time, this dataset allows us to track the diffusion of a set of related technologies over the same network.

Reflection/homophily could be controlled for in several ways: using potential exogenous variation in exposure, propensity scores, using adoption of one technology as a control for adoption of a similar technology, or limiting analysis to technologies that are directed.

Potential directions:
  a) Intervention. For example, informative text messages could be sent to selected members of a treatment group; a month later, treatment and control groups could be surveyed over the phone to see how the message spread. Or, members could be invited to participate in a puzzle that requires information to be shared and exchanged. Call data can be used as covariates or to stratify samples.
  b) Comparing methodologies. As an additional component, experimental estimation of peer effects can be compared with nonexperimental estimation.
  c) Social network validation. Mobile phone calls suggest connections between individuals, but how well does this inference match actual social networks on the ground? Predicted networks could be compared with an individual's perceived network.

References
Blumenstock, J., D. Gillick, and N. Eagle. 2010. "Who's Calling? Demographics of Mobile Phone Use in Rwanda.".
Butera, Saul. 2010. "Rwanda: Gov't Plots for Cheaper Mobile Handsets." *The New Times (Kigali)*, January 30
        http://allafrica.com/stories/201002010149.html (Accessed April 28, 2010).