Moving beyond Operations: Leveraging Big Data for Urban Planning Decisions

Steve French, Camille Barchers and Wenwen Zhang

Abstract

Big data is here: urban infrastructure systems are being instrumented to provide continuous reports on their performance; buildings are monitoring and reporting occupancy and energy use; distributed water and air quality sensors are providing real time information on dozens of environmental parameters. Cell phone location data is providing a detailed view of the activity patterns for millions of urban residents. However, when the utility of big data is discussed almost all of the examples provided are short-term management applications. There are very few examples of big data being used in long range planning. This paper discusses why big data is particularly well suited to short term management applications and identifies the factors that have limited its use for longer range planning. This paper also provides examples of how big data can be abstracted in ways that are useful for long range planning, and how these uses differ from the short-term management applications that are so commonly discussed.

S. French (Corresponding author) • C. Barchers • W. Zhang
School of City & Regional Planning, Georgia Institute of Technology, Atlanta, GA 30338, USA
Email: steve.french@coa.gatech.edu
C. Barchers
Email: cbarchers3@gatech.edu
W. Zhang
Email: wzhang300@gatech.edu

1. Urban Big Data

Traditional structured data sets can be thought of as a large cube. Like a simple flat file, big data sets can consist of a large number of rows (or observations) that are described by a large number of fields (or variables). Many big data sets add a third temporal dimension that includes recurring observations over time, sometimes on a second by second basis. Many of these data sets can be joined to variables in other structured data sets using some common identifier. Since many of these records are tagged with geolocation or a time stamp, and sometimes both, time or location can often be used to join otherwise unrelated data sets. In addition to this traditional structured data, we now have vast amounts of unstructured data (e.g. drone videos, Tweets, Facebook posts, YouTube videos, Foursquare check-ins, surveillance videos and much more). As shown in Figure 1, while unstructured, much of this new data incudes time or location information that allows it to be linked to the more structured data. The world has rapidly moved from a data poor environment to a data abundant environment.

If you live in a metropolitan area, think of all the cars on the roads in your metro area at rush hour. Assume that each of those cars contains a cell phone or two. Each of those phones is keeping track of its location on a second by second basis and reporting it back to its network (Herrera et al., 2009). Even today, Google, Waze and other mapping services are polling the location of these cell phones and using that information to produce real time traffic maps. These maps can help drivers change their routes to avoid congestion and are already improving traffic flow incrementally. Now imagine storing all that data for a year, or maybe twenty, and using it to identify recurring patterns of congestion and how drivers adapt to it.

that is an extension of today's crash avoidance and blind spot detection technologies. This urban cruise control system will manage much of the car's operation, but can be overridden by the driver, much like the automatic pilot on an aircraft.

Automobile manufacturers have already begun to develop these capabilities in cars on the road today. Mercedes Benz has developed "collision prevention assist" that uses radar to scan the environment to assess if vehicles are stopped or slowed ahead of the moving vehicle. This assist mechanism alerts drivers to take action by braking, and responds by increasing the intensity of the braking based on the distance between the car and the obstacle ahead. Additional technologies available in Mercedes vehicles include "attention assist", "active blind spot assist" and "active lane keeping assist." Other car manufacturers are employing similar technology for parking assist that allows a car to autonomously park itself using cameras, sensors and radar to detect the proximity of surrounding objects and parked cars. Ford, Toyota, Hyundai, and Land Rover currently offer these technologies in their late model cars and SUVs. Each of these assisted driving technologies can act autonomously if the driver does not respond to audio, visual, and tactile prompts from the car. These technologies represent capabilities that will be required by fully autonomous vehicles, but will first be implemented as features to assist drivers rather than replace them. These technologies are already on the road today.

To improve safety and optimize traffic flow, the cars will also communicate with each other. Typically, each car will be exchanging data with three to four dozen nearby cars about its speed, position and destination. This communication between vehicles combined with autonomous control features will allow vehicles to operate with reduced headways. By eliminating the response time required for human braking, vehicles can be spaced more closely, thereby increasing the capacity of existing highways (Dickerson, 2015). Mercedes Benz is already selling cars with "distronic plus" that uses an adaptive cruise-control feature to pace the car's speed with the flow of traffic and autonomously brake to avoid collision. Perhaps more significant than the impact these automobile technologies will have on individual drivers is the impact they could have on future infrastructure projects. It has been estimated that traffic congestion delays could be eliminated in metro Atlanta within the next ten years by utilizing technologies such as urban cruise control, mobile applications for ride-sharing, and HOV to HOT lane conversion on the fly (Dickerson, 2015).

Now combine all this structured data with the feeds from surveillance cameras, red light cameras, drones and all the posted video, image and text data, that contain time and location tags and we begin to see the magnitude of the data that is becoming available. Individual photos and videos can be knitted together to provide a comprehensive and dynamic image of the city. Combining this unstructured data with the vast amounts of structured data provided by intelligent systems and you have not just big data, but urban big data.

2. Beyond Management

But notice that the illustrative examples above are about the operation of existing systems. The systems may be able to learn and improve and even evolve over time, but few of us would be willing to surrender our ability to shape the urban environment, which will soon house 80 percent of our species, to a set of operational algorithms. Optimization approaches have long been found inadequate when it comes to designing and planning the complex interacting systems that comprise an urban area (Harris, 1999). That requires values to guide the tradeoffs required to choose among competing and conflicting priorities. Big data provides a way to build and test theories about cities and advance our ability to model the urban area and the behavior of households, firms and institutions within it. We can use urban big data to build, test and advance our theoretical framework of how cities grown and develop (Bretagnolle et al, 2006). A strong theoretical foundation is necessary to intentionally modify urban systems and settlement patterns to improve the environmental, economic and social conditions in which humans live. This requires a set of models that can help link potential interventions to intended outcomes. Big data can help us do this better with fewer unintended consequences, but it cannot be a substitute for causal models that connect specific policy interventions to outcomes.

In his widely cited "End of Theory" article Anderson (2008) argued that with the advent of big data, we can simply observe the patterns and correlations in the data and will not need to build explanatory models. For the last 200 years science has operated by developing hypotheses, constructing models, collecting data to test those models. Anderson argues that we no longer need to know why people and systems behave as they do, just that they do. No causal models are required because "with enough data, the numbers speak for themselves." Other scholars have followed this line of thinking, while considering the challenges that big data poses for researchers and

3. Urban Activity Systems

Chapin and his colleagues developed a human-centered view of the city based on human activity patterns. This approach provided a framework to describe how households, firms and institutions interact with a fixed pattern of land uses and infrastructure to meet their economic and social objectives. Members of households live, work, shop and play in the midst of a complex set of land uses that provide them with a wide variety of opportunities. A household locates itself in order to meet those needs as well as it can given its budget constraints. Similarly, firms locate their plant and equipment to balance their needs to import raw materials and ship finished products in and out, while employing workers and attracting customers to their location. In the short run these actors operate within a fixed pattern of land uses and infrastructure. But, over time the urban area will change its size, shape and form. The community can shape and re-shape the urban landscape to make it easier for actors to meet their needs. Planners and policy makers have the opportunity to make the urban landscape "more user friendly" over time through infrastructure investments and changes in the land use pattern. In Chapin's time we did not have the means to collect the data necessary to understand these complex patterns or to analyze the data if it had been available. Urban big data makes this possible.

Historically, communities have made urban planning decisions with limited information and a fairly crude understanding of how various factors interact. This has produced some successes, such as the elimination of tenements and separating residential uses from noxious industrial activities. But it also produced some unpleasant surprises, like the social isolation, increased energy use, and other environmental impacts that have resulted from the decentralized monocultures that characterize the current suburban land use pattern (Duany and Plater-Zyberg, 2001).

But now urban big data provides the opportunity to understand the nuances of these systems and to find correlation and causality that was simply not possible with occasional sampling through small surveys. We can now get a detailed view of how people, firms and institutions use urban space and more rapidly identify the complex patterns of behavior that characterize an urban area.

3.1 Travel Behavior

Metropolitan Planning Organizations (MPOs) regularly conduct household travel surveys to update their regional travel demand models. These surveys

3.2 Potential of New Sources of Travel Behavior Information

Other sources of information on travel behavior include mobile applications like Uber and Lyft. Uber is the current leading mobile app that connects riders with private drivers willing to share empty seats in their vehicles. While serving clients, Uber also collects data regarding how people travel through the city (including origin, destination, route, and time). Although Uber has recently come under fire for its privacy policies and faces legal challenges from traditional taxi companies, the ability to use this type of data for long range transportation planning is still worth exploring. This kind of data is critical for city and transportation planners to understand because it represents travel behavior of specific subgroups at a more fine grained level.

Earlier this year, Boston partnered with Uber, in the hopes that the ride-sharing service could help shine light on the city's transportation needs. In the future, Uber will provide Boston planners with their customer's trip information, including trip origin, destination, departure time, distance, and duration. However, to protect riders' privacy, the locations will be aggregated to zip code level. Given this information, planners will have access to a variety of detailed travel data that was previously unavailable; for example, planners will be able to estimate the travel time between two general locations by time of the day. Although the Uber data will have representation issues (i.e. people who use Uber may not have the same demographic and socioeconomic profile as the overall city population) planners still expect this information to be useful as a way of reflecting the vehicular traffic conditions on roads throughout the day.

Boston's chief information officer, Franklin-Hodge suggests that the Uber dataset will become one of the most helpful datasets in terms of informing their transportation and planning conversations for long-term development initiatives (Enwemeka, 2015). Many cities have also attempted to use data from Waze, a crowd-sourced map application, to obtain traffic conditions by location and time. Both Uber and Waze data are limited in terms of data resolution and population representation. So far, these datasets cannot completely replace conventional travel surveys. For example, the data available from these providers does not include important trip characteristics, such as occupancy and purpose. In this regard, the data could be used to supplement a household survey, but not yet as a replacement.

and understand the impacts of potential interventions. As shown in the activity systems case, big data can certainly help develop theory based on the collection of data that more accurately represents current behavior. The short-term immediacy of big data makes it useful and compelling for real time management applications, but to be useful for planning, this tsunami of data needs to be channeled into theoretical frameworks. In this paper, we propose using the classic conceptual model of urban activity systems to guide the use of this data for long range planning.

So how can we harness the power of big data to go beyond operations? First, we need to make these data sets available to the public agencies responsible for urban planning. Big data is too valuable to only serve the commercial interests of a small number of corporations that are positioned to collect and utilize it. If big data is to serve the public interest, it needs to be made available to the public agencies entrusted with the long range planning function. But, to adequately engage the public, the whole set of stakeholders interested in setting the goals and policies to guide urban development needs access to this data in some form. Obviously, most of these public agencies and certainly small citizen groups and non-profits need access to the data.

Access to data will also require better tools to visualize and analyze this information, especially when integrating data with new and existing urban models. Statistical methods that were useful for generalizing from small samples to larger populations are no longer appropriate tools. When you have all of the data describing a population or a system, the problem is not generalization, but data reduction and abstraction. Data analysis methods familiar to computer scientists have proven to be promising for generating understanding in a data-rich environment. These include machine learning and data visualization. Machine learning is a core subarea of artificial intelligence; it uses computer algorithms to create explanatory models. There are different types of learning approaches, including supervised, unsupervised and reinforcement learning. Although some of the technologies may be completely new to planners, the actual methods turn out to be quite familiar. For example, the regression model is one of the methods that is frequently used in supervised learning processes. Planners who work with remote sensing images often apply supervised classification methods to reclassify the images into land cover images based on various color bands in the image. However, planners may not be familiar with other machine learning methodologies or algorithms, such as unsupervised or reinforcement learning. Unsupervised learning tries to identify regularities (or clusters or groupings) in the input datasets without correct output values provided by the supervi-

along with better analytical tools and more secure privacy protections are needed to allow big data to realize its full potential for urban planning, but we are hopeful that planners, in partnership with data scientists, are ready to accept this challenge.

References

Anderson, C. (2008). The end of theory. *Wired.* 16(7), 16-07.

Atlanta Regional Commission, website: http://www.atlantaregional.com/transportation/travel-demand-model/household-travel-survey

Batty, M. (2013). *The New Science of Cities.* MIT Press.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provacations for a cultural, technological, and scholarly phenomenon. *Information, communication & society,* 15(5). 662-679.

Bretagnolle, A., Daudé, E., & Pumain, D. (2006). From theory to modelling: urban systems as complex systems. *Cybergeo: European Journal of Geography.*

Chapin, S. & Kaiser, E. (1979). *Urban land use planning.* University of Illinois Press.

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and geographic information science,* 40(2). 130-139.

Dickerson, S. (2015). Technology plus investment could eliminate congestion. Atlanta Journal Consitution. 13 April 2015.

Duany, A., Plater-Zyberk, E., & Speck, J. (2001). *Suburban nation: The rise of sprawl and the decline of the American dream.* Macmillan.

Enwemeka, Z. (2015). Boston will use Uber data to Tackle Traffic Issues. Retrieved from: http://www.wbur.org/2015/01/13/uber-boston-data-partnership

Few, S. (2009). *Now you see it: simple visualization techniques for quantitative analysis:* Analytics Press.

Garnter (2013). Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units by 2020. 12 December 2013.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography,* 3(3), 255-261.

Harris, B. (1999). Computing in planning: professional and institutional requirements. *Environment and Planning B.* 26. 321-332.

Hasan, S., Zhan, X., & Ukkusuri, S.V. (2013). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedins of the 2nd ACM SIGKDD international workshop on urban computing*. P 6. ACM.

Herrera, J.-C. ,Work, D., Ban, X., Herring, R., Jacobson, Q., & Bayen, A. (2009). Evaluation of traffic data obtained via GPS-enabled mobile phones: the *Mobile Century Field Experiment. Transportation Research C,* 18. 568-583.

Klosterman, R. (2013). Lessons learned about planning. *Journal of the American Planning Association,* 79(2) 161-169.

Lazer, D.M., Kenedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis.

Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech & Intell. Prop., 11, xxvii.*

Smiciklas, M. (2012). The power of infographics. *Using pictures to communicate and connect with your audience. Indiana: Que.*