The Use of Participatory Urban Sensing Data in Urban Infrastructure Investment Assessments: Insights from Two Delphi Surveys in Beijing

Ying Long, Steve Denman, Debbie B Deng, Xiao Rong, Xihe Jiao and Ying Jin

Abstract

In this paper we consider the usefulness, feasibility and methodology of voluntary participatory urban sensing for contributing to the assessment of alternative urban infrastructure investment plans. Participatory urban sensing in this context means the involvement of individual citizens and community groups in collecting, sifting and using urban sensing data and associated meta-data for a specific purpose or purposes. We further limit the discussion to volunteered sensing data. The review and two online Delphi surveys carried in Beijing show that voluntary participatory urban sensing could potentially fill an important gap regarding inputs into the data-hungry though necessary predictive models. The questionnaire results have provided the first parameters for survey design and implementation [The Delphi work is on-going and we will be able to report insights from a further iteration of the survey at the conference]
1. Introduction

In this paper we consider the usefulness, feasibility and methodology of voluntary participatory urban sensing for contributing to clear, open and evidence-based assessment of alternative urban infrastructure investment plans. We use the term
participatory urban sensing to mean the involvement of individual citizens and community groups in collecting, sifting and using urban sensing data and associated meta-data for a specific purpose or purposes. For reasons which will be clear in the next section, we further limit the discussion to volunteered sensing data. Voluntary participatory urban sensing is one of many ways of human-centric sensing (see reviews by Estrin, 2010; Srivastava et al, 2012). This is a fast moving field, which is so far dominated by small scale, opportunistic studies. Voluntary participation appears to be under-researched, in stark contrast to other smart data applications. Furthermore, the use of smart data in improving specific urban service operations has made remarkable progress in recent years, whereas using such data for city-scale, longer term policy and planning is rare.

The use of urban sensing data involves complex social, political and governance issues. For it to be used securely, effectively and true to the stated purpose(s), appropriate frameworks, rules and standards will be required. Frameworks for opening up data and information held by companies, institutions and government agencies are being developed in Europe through new guidelines (e.g. European Union, 2014) and coordinated policy discussions (ODUG UK, 2015). By contrast, the understanding of the potential use of voluntary sensing data contributions is lagging behind. In particular, there is little knowledge of who the willing contributors are, and what should the terms be to engage with them. This paper aims to develop an initial structure for examining the issues based on two Delphi surveys in Beijing, China where there is a real and urgent need for new data sources in infrastructure investment assessment.

2. Related literature

This research is connected to many hitherto unrelated disciplines. Because it is fast moving, we propose to structure the literature review in two generic parts: we first look at what is required for infrastructure investment assessment, and then contrast that with the opportunities and potential of voluntary participatory urban sensing. The approach still has a long way to evolve as it adapts to emerging technologies and user dynamics in the years to come.

Requirements for infrastructure investment assessment

In contrast to the short term, day to day and rapidly shifting infrastructure operations (for which there is already a growing body of sensing applications), capital-account investment in infrastructure are characterized by asset longevity, indivisibility and inertia: the majority of infrastructure investment projects are slow to develop and may need to be progressed in many phases – the projects give rise to long lived structures which usually have asset lives of 30-60 years or longer; the build out may be slow and uncertain for it is subject to ever changing political and financial cycles.
as well as challenges in construction. Once built, the configuration of the infrastructure can rarely be changed without major retrofit costs. In other words, the challenge in decision making arises because the level of demand for a given piece of infrastructure may be uncertain or even unknowable at the time of initial design.

The part of the city that it interacts with may take years or even decades to adapt to a new piece of infrastructure. For instance, housing, shops and related facilities in west London suburbs did eventually follow the expansion of the suburban rail and metro lines as the Victorian rail investors wanted, but their construction lagged some twenty years behind the railways and the neighbourhoods took even longer to mature (Stanilov and Jin, 2014).

The most uncertain factors that affect the demand of a major piece of infrastructure tend to transcend geographic boundaries. In part this is because its long term service catchment is wide, but more fundamentally, this is because the behaviors of infrastructure users are conditioned by their individual circumstances, and the patterns emerging at the system-level are driven by complex interactions of disparate decisions across the mega-city region and beyond.

The current debate over declining trips rates and car miles driven across a number of rich countries has attracted the attention of the road traffic analysts (Goodwin, 2012; Le Vine and Jones, 2012; Volpe Transportation Center, 2013). The average number of trips made by the urban population appears to be decreasing over the last decade, and the average number of car miles driven has also been falling. Are those changes a temporary blip in the data or a break in past trend? If the latter they could have a major implication for new roads, bridges and even public transport projects. To understand the implications of such phenomena, new evidence needs to collected from a wide range of sources including social and economic research, transport surveys and econometrics in order to gain a rounded view of the trends in road use, the drivers behind them and the extent of the uncertainties (UK DfT, 2015).

To address the analytical needs for justifying such investments, a dynamic or quasi-dynamic urban modelling framework that can foresee at least a couple of decades is now considered more suitable (see Anas, 2007; Bröcker and Korzhenevych, 2011; Simmonds et al, 2013; Jin et al, 2013).

For urban sensing, the analytical needs for infrastructure investment therefore imply that the data will need to be collected (1) as a time series that shows the evolution over many years if not decades in order to reveal any trend breaks, (2) for a wider geographical area which is defined by the expected long term service catchment of the infrastructure facility in question, and (3) to have a clear understanding of who the users are, how their behaviors are related to their circumstances, the motivations behind the urban activities, and the system-level emergence that arise from interactions of a large number of disparate decisions.
Unsurprisingly, the main sources of data and knowledge that inform the investment assessment analysis have traditionally come from dedicated surveys using individual and focus group interviews, where the survey initiators have reasonable control regarding the scope, coverage and timing of data collection; to understand the circumstances and motivations, they would also need an in-depth understanding of the personal profiles of the interviewees (though not necessarily their identity). If such data is continuously collected its value will obviously grow.

The downside to such information gathering is also obvious: they tend to be so expensive that even the very rich countries can only afford relatively small, bespoke datasets which severely limit the geographic and temporal granularity. The lack of data in turn holds back the availability and quality of coherent predictive models which are essential for a proper analysis and comparison of alternative investment options.

Data that has been collected for more general purposes may have greater granularity, such as the Population Census or government administrative databases for companies and properties. Such general purpose data is also costly to collect and maintain. Their extensive coverage and granularity often require substantial time for processing, which could cause a long time lag between data collection and release.

However, a particularly relevant issue is the need for disclosure control in the general purpose datasets. By definition, disclosure control must remove any part of the data that is considered sensitive by all providers before releasing for general use. Careful disclosure control is vital for sustained data collection and use and we absolutely support that principle. However, applying common criteria that satisfy all providers nevertheless could greatly diminish the potential value of the data for analysis and modelling regarding drivers for behaviour. Often some of the data providers may be willing to share their data to a greater extent – especially for specific uses that they endorse - which could provide additional valuable insights that are not available currently.

Setting up new voluntary agreements as part of established general purpose databases is at best unwieldy and most likely to be impossible. It would seem that it is worth examining alternative ways to engage with those who are willing to volunteer data to a greater extent and in greater depth. Of course, for whichever approach it must be done under robust safeguards regarding privacy and data use.

In summary, the data and insights required by infrastructure investment assessment tend to be a great deal more demanding in terms of temporal and geographic coverage than for improving infrastructure service operations. Furthermore, the analysis and modelling necessary for supporting investment decisions tend to require
building up a coherent picture of not only what activities are observed, but also the circumstances and motivations that drive them. To address the shortfall in data, analysts and modellers have resorted to in-depth and targeted interviews including focus groups. However, practically all of the tried-and-tested traditional data collection methods have long term affordability issues even in rich countries, let alone the emerging economies. The resources constraints result in low geographic granularity and often long waits between repeated surveys (if the surveys are repeated at all). This means that the scope for scaling up existing datasets is difficult. The question is, can the new technics and methods of urban sensing through smart data make a step change?

**Participatory sensing: opportunities and barriers**

The last decade has seen a surge in sensing applications characterized by distributed collection of data by either self-selected or recruited participants for the purpose of monitoring personal activities, sharing local conditions and mapping physical/social phenomena. This follows an exponential spread of a wide variety of sensors, which offer opportunities for quasi-ubiquitous and real-time data-sharing.

Smart phones have been embedded with an increasing range of sensors; they also have access to quasi-ubiquitous network connectivity and social networking sites for information sharing. In addition, radio-frequency identification (RFID) tags, smart power metres, smart cameras, in-vehicle global positioning system (GPS) devices, accelerometer-enhanced entertainment devices (e.g. Wii, X-Box, Sony PS4) and activity monitoring sportswear (e.g. Nike+iPod) have all been spreading in cities. Wrist-wearable health and biometric monitoring sensors are being marketed in a big way.

The technological advances offer unprecedented opportunities for data collection at a massive scale. In most city regions the near-saturation levels of mobile ownership have perhaps for the first time made data collection through mobile phone sensing a realistic proposition for all urban locations. How to involve citizens in sensory data collection and in the associated decision-making loop has become a critical research challenge, including the management of incentives, recruitment, privacy, trust, data accuracy, system modelling and interpretation of social-sensing dynamics (Srivastava et al, 2012).

Participatory sensing refers to the vision of distributed data collection and analysis at the personal, urban, and global scales, in which participants make key decisions about what, where, and when to sense (Burke et al, 2006). Mun et al (2009) point out that the mere existence of those devices and capabilities does not make them immediately usable or scalable, and there is a need to develop integrative platforms that unite the mobile handset based data collection and the server-side processing stages to interpret such data. They use such a platform to infer the CO2 emission and health impacts arising from people’s travel on different means of transport, and
their potential exposures to health hazards such as airborne particulate matter emissions, smog, and the time integral of proximity to fast-food eating establishments.

Bisdikian (2014) further points out that the emergence of large scale, distributed, sensor-enabled, machine-to-machine pervasive applications necessitates engaging with the people who actually collect the information in order to make correct inferences. This is because the actual sensing are highly fluid activities, involving a diverse range of trust levels and intentions among the information providers and users. They propose that information transformation, such as a controlled degree of obfuscation be used to avoid inferences being made through abuses of the information collected while generating benefits to the intended information users. They discuss how quality, value and risk of information relate in collaborative and adversarial systems, and then consider the option of setting up an inference firewall in controlling and regulating the information uses. Voluntary participatory sensing can be seen as one particular type of the collaborative systems.

The individuals participating sensing activities play three general roles in human-centric sensing (Srivastava et al, 2012). First, the individuals may be the main targets of sensing. It is intuitive that such sensing would be more productive where the individuals are knowing and willing, as opposed to being involuntary or even adversarial (such as, at an extreme, when tracking enemy soldiers in a battlefield). Second, the individuals may be the main operators of sensing. This is beneficial not only for the general agility and common sense as operators, but also for their reactions to unexpected circumstances where pre-programmed procedures fail. Third, the individuals may provide additional data sources, either through annotating and commenting on sensor data, analysing and fusing data, making inferences based on their unique aptitudes and feelings, bringing knowledge and assessment, and derive useful knowledge despite the gaps and imperfections in the sensor data.

Of course the increased human involvement in the sensing process could also lead to new challenges, such as introduction of additional errors, biases, wilful obfuscation and falsification. This would imply an increased workload for participant recruitment and coordination. Furthermore, there may be new, emergent dynamics that arise from social sensing and interactions among the participants. While a significant amount of research has been undertaken along these fronts, much remains unsolved (Srivastava et al, 2012).

Nevertheless, small scale projects using volunteered and participatory data have been emerging in the field of infrastructure and urban development, building incrementally on both existing participatory planning practices and opportunistic sensing initiatives.
For instance, Steiniger et al (2012) report PlanYourPlace, a new combined social network and Geographic Information System (GIS) platform they have developed for engaging a broader range of citizens in the local planning in and around the City of Calgary. This platform is complementary to the current participatory planning practices, and can allow citizens to voice their opinions and discuss future urban development scenarios. Whilst the social networks enable participants to learn, discuss, vote and share, the GIS allows them to create plans and performing impact assessments.

The NoiseTube participatory sensing framework for monitoring ambient noise was developed by D’Hondt et al (2013) in the city of Antwerp, where participatory sensing offers the opportunities to collect noise measurement data which they believe would in principle achieve high measurement granularity in space and time. The work so far shows that participatory techniques, when implemented properly, can achieve the same accuracy as standard noise mapping techniques for a one square km test area.

The work of Manzoor et al (2013) examines the possible downsides of a participatory sensing campaign, which they believe could suffer from low-quality, misleading, or conflicting data. They think that the level of trust, i.e. the knowledge of the participants’ competence to collect data is vital for an effective sensing campaign and the overall success of this type of applications.

In summary, it seems clear that the new sensors, particularly those equipped in smart phones will revolutionize urban sensing. However, applications that benefit long range impact assessment of major urban infrastructure and urban development projects are relatively rare. By the nature of capital-account investment assessments, there is a need for more in-depth understanding of the full range of circumstances and motivations underlying the behaviours of the infrastructure users than common for improving infrastructure operations.

This means that the standard obfuscation techniques such as applied currently to human-centric sensing for preventing privacy problems and data abuse would not work well in our context. On the other hand, the emerging sensing scene suggests that the most feasible sensing activities are necessarily opportunistic, partial, time-dependent and ephemeral.

This suggests that it is necessary to find new ways of working with the human-centric sensing data, especially how the partial and ephemeral data reported by individuals could be assembled for building up a rounded picture of the behavioural patterns cogent for infrastructure investment assessments. Having a large group of volunteers who can provide detailed information about their personal circumstances and motivations along with sensing data of their activity patterns is an important component of this data collection. The data assembled this way will need to have
proper safeguards in terms of their ethical use as agreed with the contributors.

In contrast to traditional survey and interview data which are properly structured in terms of random sampling, volunteered sensing data is a biased sample, where the extent of biases cannot be controlled by the survey designers. Such biases are common among all types of human-centric sensing. However, there may be a greater scope for voluntary participation to disclose more information about the individuals, which can help survey designers understand the precise nature of the biases and work with them. The use of the voluntary participatory sensing data in a comprehensive urban modelling framework may also help to identify the data gaps among segments of the urban population. The model may be used both as a hypothesis generator and a platform for combining urban sensing data with the traditional data sources.

In our view, the immediate challenge for voluntary participatory sensing is that we know little about who the volunteers are and the terms that should be used to engage with them successfully and ethically. To make progress we have carried out two online pilot surveys, which in themselves are a rudimentary exercise of voluntary participation. We report the surveys and their results below.

3. Participatory urban sensing Delphi questionnaires

The two questionnaire surveys were carried out in the Delphi format, in the sense that the respondents are treated as knowledgeable individuals who are not only asked factual questions but also those involving subjective judgement. The format includes further iterations i.e. the respondents are shown the overall results from the questionnaire survey and have a chance to modify their judgement and provide new comments. [We will report the results of a new iteration at the conference.] However, the questionnaires are posted on the webpages (see below) rather than delivered to specific recipients, which is a key difference from the traditional Delphi surveys.

The two questionnaires were first released on SOJUMP (http://www.sojump.com/), a professional survey website on December 5, 2014. The first (Questionnaire A) targets university students and young researchers, and the second (Questionnaire B) senior researchers and team leaders. In order to draw attention from the targeted respondents pools, the webpages of the two questionnaires were posted on Sina Weibo by Beijing City Lab. The news of the questionnaires was also disseminated through Wechat, a widely used mobile text and messaging service by the social media networks.

The prospective respondents were informed that the questionnaires were for studying the use of participatory urban sensing in contributing to the acceleration in urban air quality improvements in Chinese cities. They were invited to answer 12
sets of short questions through an easy to operate webpage. The questionnaire is semi-structured – the respondents may ignore unwanted questions and volunteer additional comments.

Because the concept of participatory urban sensing is little known in China, the number of respondents is small relative to our estimated size of potential respondents pools [Ying L to fill in the estimated pool sizes, relative to the numbers of respondents we have]. This in part is an information dissemination issue which is being helped by online surveys such as what we were carrying out. However, the actual respondents appear to have balanced gender, age, education and work profiles in line with the target population (see Figure 1). In addition to the specific questions of their personal profiles, we have also asked whether the respondents would wish to provide email contacts to be acknowledged specifically and to help the researchers further; all but two respondents responded positively to the first query and all responded positively to the second query.
Figure 1: Summary profiles of survey respondents – students and young researchers on the left column and senior team leaders on the right.
4. Responses from two Delphi groups

We summarise the responses from the two Delphi groups below through answers to specific questions. Note that there is a key difference between the questionnaire for the students/young researchers and that for the senior team leaders: since the team leaders are expected to delegate the urban sensing field work to their teams, we ask the team leaders only for their perceptions of their teams for many of the questions below. Note the team leader’s perceptions apply to their whole teams (students/young researchers). Because only a self-selected group of the students/young researchers have responded to our survey, the differences between their answers and the perceptions of the team leaders are not directly comparable – they should be read instead with the above caveat in mind.

![Figure 2: Knowledge of participatory urban sensing is limited – 75% of the students and young researchers have not heard of it](image)

The start of the questionnaire explains in clear and generic terms what participatory urban sensing is. The respondents are then asked whether they have heard of the term. It is clear that the vast majority has not heard of the term – this is even true with the team leader group (Figure 2).
There is tremendous enthusiasm for participatory urban sensing, once the respondents realise what it does (Figure 3). Of particular note is the fact that the majority from both groups of respondents are willing to be involved in analysis, result interpretation and data collection design besides data gathering. This could imply a major resource that is little tapped in sensing activities so far.

Figure 3: There is enthusiasm for analysis, interpretation and data collection design besides data gathering

Figure 4: Nearly 80% of respondents have previous experience with sensing apps, with fitness monitoring being by far the most familiar
Interestingly, both groups have most experience with fitness sensing apps (Figure 4); whilst this is understandable for the students/young researcher group, it is puzzling with the team leaders group which may need further investigation. Smart phone or tablet is the most preferred sensing device (Figure 5), and if they have to carry a special instrument, it should be no larger than an iPhone (Figure 6).

**Figure 5:** Mobile phones are clearly the preferred device for monitoring

**Figure 6:** An iPhone-sized monitoring device is the most popular, even more than a wrist e-watch
Figure 7: Any requirements for user intervention in the course of sensing will at least halve the number of volunteers.

Figure 7 and Figure 8 indicate two major challenges in the design of sensing operations: whilst the respondents may be willing to provide personal and circumstantial information generally, they do not like the idea of doing so whilst they are carrying out the sensing – the integrative platforms that can successfully infer the contexts of sensing may well prove to be the most suitable approach even for voluntary participatory sensing (see e.g. Mun et al, 2009).

Figure 8: Short durations are much preferred – best done for 20 minutes although 10% of students/young researchers.

<table>
<thead>
<tr>
<th>Do you wish to participate in ...</th>
<th>Students/Young researchers</th>
<th>Perception of team leaders</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 hr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 hrs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>five working days of the week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>all seven days of the week</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
researchers would do for as long as one month.

Nevertheless, there does not seem to be a large difference between sensing at a time required by a given research project and doing that at a time that is convenient to the person carrying out the task (Figure 9). Perhaps the students and young researchers in Beijing have become used to the local format of temporary paid work, and are thinking along those lines when responding (Figure 10 and Figure 11).

Figure 9: Sensing when required vs sensing when it is convenient: the willingness is not that different

Figure 10: 70% would expects pay among the students and young researchers

Figure 11: Percentage of students and young researchers who believe that monitoring should be paid work
Figure 11  The level of expected pay centers upon 500 yuan/week

Figure 12: More than 95% happy to respond to queries but only two thirds to redo data collection

Figure 12 reports the really good news that the overwhelming majority is able to respond to researchers’ queries and a reasonable number of respondents could redo the data collection exercise if required. There is a healthy desire to compare with sensing results from other contributors (Figure 13). However, only around two
thirds are interested in following up the research outputs and continuing discussions (Figure 14). Note those proportions are generally consistent with the responses reported in Figure 3 regarding the involvements in results interpretation. This could imply that specific engagement may be more valuable with data analysis. Although all of them have left email addresses for future contacts, a small group of respondents have not given permissions for reusing the data in similar research projects (Figure 15). This highlights the importance of the management of the dataset – e.g. the capability of separating permissible and non-permissible responses - particularly where there is a large and viable dataset for reuse like in our surveys.

![Figure 13: There is overwhelming curiosity (c90%) about the results from other contributors](image1)

![Figure 14: Less than two-thirds are interested in continuing discussions](image2)
5. Conclusions

The review and discussions in this paper show that voluntary participatory urban sensing could potentially fill an important gap regarding inputs into the data-hungry though necessary predictive models for assessing alternative infrastructure investment plans.

The questionnaire results have provided the first parameters for survey design and implementation in the context of Beijing: the current knowledge regarding participatory urban sensing is very limited – 75% of the students and young researchers have not heard of the concept, which could severely limit the scale of initial applications; there is however great enthusiasm, which extends to data analysis, result interpretation and data collection design besides data gathering; nearly 80% of respondents have previous experience with sensing apps, with fitness monitoring is being by far the most familiar; mobile phones are clearly the preferred device for monitoring, and if they have to carry a special instrument, the device should not be larger than an iPhone; it seems that requirements for user intervention in the course of sensing will at least halve the number of volunteers; short durations are much preferred – ideally no more than 20 minutes although 10% of students and young researchers could do for as long as one month; the willingness is not so significantly affected by when the sensing is to be carried out – sensing required would only depress the number of volunteers by 15% or so when compared with sensing at a personally convenient time; nevertheless 70% would expect pay (of
around 500 yuan a week) among the students and young researchers - this could be a local matter in Beijing as students and young researchers commonly carry out additional tasks with supplementary pay; more than 95% are willing to respond to queries but only two thirds to redo data collection; there is overwhelming curiosity (around 90%) about the results from other sensing contributors, although only less than two-thirds are interested in continuing discussions with the research project team; finally it is worth noting that a small proportion (under 10%) who are not willing to give permission for reuse of the responses beyond the stated purpose, even for similar research projects which highlights the need to factor this into appropriate management of the data.

The survey results have some obvious limitations. So far we have only been able to capture the attention of a small group of enthusiasts. This is mainly because the concept of participatory sensing is still not well known among the potential pool of volunteers, and we hope the surveys of this kind can help raise the awareness. Many respondents commented that they found the concept new and inspiring, which suggests that further publicity would help. Secondly, the sample is self-selected, and all the percentages above relate to this segment of the population. Thirdly, both technology and user dynamics are rapidly evolving. The fast changing nature of this subject suggests that we should develop the methodology with the volunteer group we have and gain experience of the contributor dynamics. In this context, the Delphi survey method is a useful one to facilitate the interactions with the respondents, engage with them, and shape up the sensing methodology accordingly.

6. References


Estrin, D (2010). Participatory sensing: applications and architecture. IEEE Internet
Computing, IEEE, 14:1


Stanilov, K and Y Jin (2014). Exploring the relationship between infrastructure and land redevelopment: an empirical analysis. ITRC Conference on The future of
national infrastructure systems & economic prosperity, Cambridge.

Stefan STEINIGER, Mohammad Ebrahim POORAZIZI, Coral A.M. BLISS-TAYLOR, Ehsan MOHAMMADI and Andrew J.S. HUNTER, (2012). PlanYourPlace: Merging Social Networks and Participatory GIS for Participatory Planning. FIG Working Week 2012 Knowing to manage the territory, protect the environment, evaluate the cultural heritage Rome, Italy, 6-10 May

