Chapter 1
The "Human Sensor:" Bridging Between Human Data and Services

Neal Lathia

Abstract

Data from 'human sensors' is increasingly easy to collect. Yet how may systems be designed that put it to use? This chapter discusses this question in three steps. First, we describe how the increasing ubiquity of digital systems is facilitating the creation of streams of human data. We characterise these data sources according to their purpose, obtrusiveness, structure, and hierarchy. Then, we address the kinds of systems that are already reaping the benefits of these data sources; they are broadly categorised as recommendation, retrieval, and behaviour-mediating systems. Finally, we describe a case study of potential systems that may be built to support urban travellers by leveraging the data that travellers themselves create while navigating their city. The chapter concludes with three open research challenges, related to understanding the context of data creation, the systems that are designed to use this data, and how to best architect a bridge between the two.

1.1 Introduction

The technology that we encounter in our daily lives is increasingly characterised by its ability to collect from, store, and process growing amounts of diverse data streams about human life. These sources of data range from small to large and offer varying levels of insight into our behaviours, thoughts, decisions, and actions. Researchers' ability to source from these data streams gives rise to the idea of *humans as sensors*: existing and technologically feasible (social, sensing, smartphone, or otherwise) systems digitise a growing number of facets of life, and provide access

Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
neal.lathia@cl.cam.ac.uk

to data that historically remained quantitatively unavailable.

While the existence of human data sources continues to grow, an ongoing challenge that researchers face is how to go beyond analysis, and towards designing and building systems that put these datasets to use. The overarching question is: if we can source data from humans (who, in doing so, act as our 'sensors'), how do we fit these people into the design or broader framework of systems that leverage the data's value?

In this chapter, we consider how researchers may reason about building the bridge between human data and systems. We do so by discussing three broad questions. First, where does human-sensor data come from? The literature in this area is diverse and fragmented, yet crowd-sourcing, participatory sensing, and database mining all share remarkable similarities in terms of the data that they provide. We therefore characterise these data sources in terms of their original purpose, the obtrusiveness of collecting them, and their underlying structure. Second, what system models exist where this data is applicable? In particular, we discuss recommendation, retrieval, and emerging behaviour-mediating technologies; each of these provide a framework for applying human data to serve different needs. In doing so, we expose a set of examples of systems that have been built using data from crowds and sensors. We highlight how these system models are applied to human mobility data by describing a case study of using public transport access records to build automated alert and fare-recommendation systems. Finally, we discuss three open research questions in this space, which encompass the spectrum of system building: understanding data representativity, uncovering how system design affects learning about humans, and the challenge of reasoning about relevance and effectiveness across different application domains.

1.2 Generating Data Streams

We begin by discussing the context of using humans as sensors. Broadly speaking, a *sensor* is a means of translating a physical phenomenon into a digital signal; for example, a gyroscope measures its orientation. Referring to a "human sensor" acts as an umbrella term for our growing ability to capture data streams about human life—activities, movements, thoughts, behaviours—and encompasses the variety of both explicitly and implicitly available means of collecting data from or about people, via digital systems.

For example, the widespread adoption of smartphones (as well as other sensor-enhanced devices) now means that people regularly carry items that are instrumented with means for collecting data streams that tell us about their behaviour; moreover, smartphones are, naturally, built as interactive devices and thus ideal for collecting data that is manually input while on the go. Researchers can therefore leverage these devices to collect data via both sensors [17] and participant’s direct input [27].

How does all this data collection come to fruition? The research literature that has begun to emerge in this space often discusses contributions as broadly related to *participatory sensing* or *crowd-sourcing*. The narrative behind the former [8] is that the increasing ubiquity of sensor-enhanced devices gives rise to the potential of collecting data from a community of, for example, participating smartphones. The focus here tends to be on the availability of sensors [30]: the spread of smartphones throughout the world has translated the historical problem of *deploying* sensor networks into one of *harnessing* volunteers’ sensors for data collection. Crowd-sourcing [16], instead, focuses on the opportunity that arises to tackle large problems by means of groups of volunteers, where each individual’s contribution may be small (e.g., writing an article or, indeed, appending a couple of sentence to one) but the group’s output is significant (e.g., creating Wikipedia). Much like participatory sensing, the key factor that determines the success of crowd-sourced data collection is the ability to harness, engage, and maintain a community of contributors. Delineating the nuances between these two groups therefore goes beyond the scope of this chapter: instead, we focus on similarities and where the two methods are beginning to meet. Both crowd-sourcing and participatory sensing are active means of generating data: advances in mobile devices now means that data from “crowds” is often accompanied by sensor data (e.g., geo-located tweets [43]), crowds’ smartphones can be used to work on sensor-related tasks [61], and sensor data can be harnessed by means of social systems (e.g., fine-grained GPS mapping from Foursquare check-ins [56] and merging social activity sharing with sensor sampling [26]). In fact, some research now blends the terms together into *crowd-sensing* [46].

An alternative source of human data arises, instead, from those repositories of information that are automatically created as we use different systems: many of these databases have presented the opportunity for research that comes as a consequence of the pre-existence of data, unlike the more traditional research that solicits defining questions and hypotheses prior to data collection. Increasingly, a variety of daily tasks that we engage in create digital “footprints” including: renting a shared bicycle [22], making a telephone call [50], clicking on web links [48], taking a photograph [23], and entering a public transport system [7] are amongst a growing family of actions that are automatically logged. More importantly, while individual entries may, at face value, seem meaningless, the aggregation of large samples of these datasets convert them into invaluable resources for insight into human life, even though they were originally created to serve other purposes.

The landscape of human-driven data that is available to modern-day researchers is, therefore, seemingly limitless. How can it be characterised? In the following section, we review how many of these data sources are being put to use; we close here by broadly characterising the varying qualities of human-sourced data:

1. **Obtrusiveness**. A hallmark characteristic that differentiates different sources of "human sensor" data is the effort and commitment that is required by participants in order to serve the researcher's purposes. Consider, for example, the difference between participatory sensing to gauge transportation modalities [51] vs. sensor-augmented experience sampling to measure the geography of happiness [39]. In the former, all that is required is for participants to contribute samples from their smartphone sensors, in the latter, participants must manually complete momentary mood assessments.
2. **Original Purpose**. Human-sourced data is further differentiated by considering why it was originally created. For example, the original intent behind tweets and Foursquare check-ins is to participate in those services' social functions by, for example, sharing your location with your friends. The fact that these data sources are now used to study mobility [41] and mood [44] is divorced from the data's original purpose; on the other hand, sensor samples collected after manually inputting an activity [26] and location traces from participants' cars [21] construct datasets that directly respond to a research question at hand.
3. **Perspective Hierarchy**. A common theme amongst human data sets is that, by being sourced from individuals, they contain a hierarchy of perspectives. For example, tweets may be used to study and build for individuals [24] or cities [44]. Similarly, sensors can reflect on both individuals' [11] or city-wide [31] behaviours. Navigating and building from these data sources often means picking one level of this hierarchy for analysis.
4. **Structural or Itemised**. A final means of characterising human-sourced data comes from asking how each source encodes human behaviour. Drawing from the previous example [26], some systems directly associate behaviour with data; in this case, sensor streams are 'labelled' with the user's current activity. On the other hand, alternative sources provide a means of analysing behaviour via the *relationships* that emerge within the data. For example, web clicks on search results encode an intent for information [48], and mobility traces encode underlying communities and social relationships [6].

In the above, we introduced the first step required to build with "human sensors:" using humans to source data about daily life. In the following section, we review how these sources of data are being translated into new insight—where empirical measurement has historically been elusive—and how the web has become the primary example of using this data as a foundation for information services.

1.3 Putting Data Streams to Use

As introduced above, the state-of-the-art facilitates collecting data from humans more so than putting that data to use within systems designed to include facets of human computation. Arguably, while data collection methodologies may suffice for scientific enquiry, the value of human data has not been fully reaped until it is part of an ecosystem that supports those behaviours that it was measured from. In this section, we discuss a number of examples where visions of how human data can be integrated into systems have appeared. In particular, we consider scenarios where the human data is not simply used as a body of knowledge that can be accessed (e.g., Wikipedia), but fully enables the existence of new systems. Broadly speaking, we decompose these systems into three categories: those that support *recommendation*, *information retrieval*, and *behaviour change*.

1.3.1 Recommending Items

The idea of giving users personalised, automated recommendations pervades the online world. Hallmark examples of these systems include Amazon.com’s product recommendation [38] and Netflix’s movie recommender [1] systems; however, social [24] and search [15] systems are now also characterised by personalisation algorithms. These systems are grounded in a view of the world where the number of available ‘items’ (e.g., movies, e-commerce products) far outnumber each user’s ability to sort through, evaluate, and find the subset that best matches their preferences [52]. These systems therefore tend to use *collaborative filtering*: an algorithmic approach that takes as input a sparse set of ratings or representations of preference, and produces personalised rankings for each user [55]. In doing so, they close the loop between human computation (since humans provide ratings that ‘evaluate’ items) and machine learning (which predict values for unrated items) to serve behavioural interests (finding new items).

The value of the model underpinning the design of recommender systems is that they are fully agnostic of *what* is being recommended. In principle, this means that *any* scenario that can be described as a set of ‘items,’ a set of ‘users’ with preferences, and required a mapping between the two may suit systems that leverage both humans’ relevance judgements and machine learning.

A growing set of examples show how this model is being applied outside of the domain of the web. A notable example relates to discovering places and social events in the physical world: for example, cell phone data sourced from a city can be used to infer those social events that people are attending, and provide them with recommendations about others that they may be interested in [45]. Similarly, GPS data from people’s smartphones can be used to find venues of interest [58, 56], or even recommend where to go after visiting their current location [42]. Beyond location-discovery, data from human mobility has been applied to, for example, helping taxi drivers find their next fare [63]. Similarly, the photos that we take have been shown to uncover the world’s interesting locations [12]: this data fits into recommender-style applications by, for example, supporting tourists who are navigating an unfamiliar place [23].

1.3.2 Mediating Behaviour

While recommendation systems seek to support those contexts where users are navigating large ‘item’ repositories, there are many domains beyond the user-item model where behaviours may nonetheless be mediated by the data that they produce. These systems are often referred to as *persuasive* [19] or behaviour-change [25] technologies, which merge the data that can be sensed or collected about human behaviour with behavioural theory about how habits are formed, changed, or maintained. In many instances, the loop between data collection and system design is closed by using the data to provide feedback to a user.

Such systems have already been applied to a host of domains. For example, users’ own data can be fed back to them in the context of sustainable travel choices [20] and physical activity [11]. In these cases, the data that is collected by directly measuring users’ behaviours is then returned via feedback interfaces which may nudge participants’ choices [28]. We further note that–rather than using the data to provide feedback to users–data about human activities can also be used to mediate the control of systems: for example, room occupancy prediction can be used to automatically tailor the management of household heating systems [29].

The key design issue in these domains seems to revolve around how to best implement behavioural theory in systems and, in parallel, how systems may augment the potential to extend behavioural theory [25]. In these cases, individuals’ data plays two roles: first, it provides a source for self-measurement and understanding [37]. Moreover, it supports automating the extent that people can track their commitment and progress towards accomplishing their goals, representing their identities, and uncovering their own inconsistent behaviours [10].

1.3.3 Monitoring and Retrieving Knowledge

A final application scenario for ‘human sensor’ data is to use the information that is collected to support retrieval-type contexts which were previously inaccessible to users. The application scenarios here are far-reaching; unlike the above, the model here relies on supporting information needs that can be expressed via a query of some type. For example, participatory sensing indicates that systems that what once relied on time tables to provide public transport information [18] can now use passengers’ smartphone sensors to crowd-source bus inter-arrival times [64]: the answer to the question “how long do I have to wait at this bus stop?” can be driven by sensor data, rather than time tables.

Beyond individuals’ mobility within a city, the data available from movements of crowds has led researchers to advocate for using these sources to guide future urban planning and leadership [57, 49]: city “leaders and service providers are looking to base decisions on data" [2]. In these cases, crowd-sourced data replaces the kinds of queries (for example, "how do people use these urban spaces?") that were previously answered by arduous field studies. For example, consider the challenge of defining urban neighbourhoods: the data sourced from people's movements now supports dynamically defining and visualising communities [13]. It is worth noting that, in some cases, similar datasets may be applicable to multiple domains. For example, the previous section mentioned using sensor-enhanced taxis to help drivers find their next fare. Similar placements of sensors on taxis can help local governments track their city's pollution [62].

Even the behaviours that accompany retrieval-systems can themselves become informative. For example, search query behaviour related to mental health has been shown to be seasonal [3]; similarly, queries about medication reveals the side-effects of combined drug usage [60]. Such data, that emerges from people using a search engine, could thus become the foundation for monitoring and retrieval tools that support medical practitioners' work.

The examples above decompose the usage of human-sourced data into three generic application scenarios: those that support recommendation, mediate (and aide to change) behaviours, and those that create new forms of retrieval systems. All of these application scenarios share a common vision: that is, taking the collection of "human sensor" data beyond mere analysis, and using it as a foundation for building systems. As above, in many of these cases the data that was collected was not originally intended to support the design of such systems. In the following section, we consider a particular use case: translating data that was originally intended to log public transport financial transactions into a basis for personalised transport information systems.

1.4 A Case Study: Computation with Smart Cards

In this section, we focus on one particular use case: turning transport smart cards into sources of information for travel services, with a particular focus on how it may be implemented in London, England [36]. Smart cards are increasingly being adopted by public transport authorities across the world: they are typically personal RFID-enabled cards that store passengers' fares or tickets. In doing so, they facilitate the process of paying for and accessing public transport and remove the need to carry paper-based tickets. However, a consequence of automating the billing process is that detailed records about millions of passengers' movements and fare purchases throughout a public transport network are created.

For example, the London public transport system uses the Oyster card: a personal contact-less smart card that allows passengers to access all of the city's multi-modal transport systems, which includes underground trains (11 interconnected lines with

270 stations), overground trains (5 lines with 78 stations), and buses (about 8,000 buses service 19,000 stops). The Oyster card itself is used to store fares, which come as both credit/pay-per-journey or travel passes, and is then used to enter and exit train stations and when boarding buses. By 2009, this system accounted for approximately 80% of all public transport trips in the city [59].

What do these records tell us about the city and its public transport passengers? Recent work has uncovered that this data contains a hierarchy of information, that ranges from patterns about individuals and communities, to city-wide behaviours: navigating between these levels demonstrates the granularity of analysis that this data enables. At the grandest scale, Oyster data reflects the overriding week-day commuting pattern of the city and shows that the metropolitan area of London, when considered based on passenger flows, has a polycentric structure [53]. Similarly, analysis of the large-scale features of the data shows how passengers travel choices relate to the financial behavioural incentives that are delineated by the transport authority [31]; for example, peak-fares seem to guide travellers' fare purchase decisions (rather than their choice to travel), students do not buy fares that they would be eligible to purchase at a discount, and the availability of free travel radically alters peoples' likelihood of taking a bus. By stepping down to the community-level, and using the data about where humans travel between to model how communities interact with one another, the same records show that mobility patterns correlate with social deprivation [35]. Finally, individual patterns of mobility measured via these smart cards uncovers the variance between different individuals' travel choices [33] and the extent that passengers overspend on public transport by failing to relate their travel behaviours to the fare most suited to them [32].

How can this researchers 'close the loop,' and turn this insight into human-data driven systems? We consider two examples, which leverage the latter analytic results: namely, that smart card data provides a means for measuring differences between *individuals'* behaviours. Therefore, a first step into this domain could entail diversifying the output of transport information services, in order to cater for personal differences. Historically, public transport information systems have been centred on the system itself (by providing, for example, the location of a transit service or, scheduled and estimated arrival times of trains or buses) and has not automatically tailored its output to individual travellers. Consider, for example, the Transport for London Travel Alerts[1]; this system requires passengers to manually set their travel choices and times. Replacing this manual input with the automated smart card data would not only alleviate users from this task, but allows for data that can predict travel times more accurately than time tables [36] and automatically rank the importance of station alerts in such a way to even capture the importance of places that travellers have not historically visited [33].

Similarly, the individual-level analysis uncovered that passengers often make the

---

[1] http://alerts.tfl.gov.uk/

incorrect fare choices when making purchases: at a city-wide level, this overspending was estimated to be approximately GBP 200 million per year [32]. Part of the problem emerges from the difficulty that people have in (a) estimating their own travel needs, and (b) linking their own forecasts with the optimal fare, particularly since the nuances between fares may not be apparent. Smart cards, however, act as an implicit diary for all public transport usage, and reveal that mobility is consistent enough that simple, moving-average based techniques are sufficient for accurately predicting those features of mobility that are relevant to fare purchases. Moreover, supervised learning techniques were then shown to be able to accurately predict between 77% (Naive Bayes) to 98% (Decision Trees) of the 'cheapest' fares that passengers need.

Both of these results demonstrate how data that has been sourced from human behaviour can be used and contribute back towards guiding it. By borrowing techniques that have been widely applied in the online world (personalisation and recommender systems) and applying them to pervasive data (from smart cards), this research demonstrates how many of the tools, techniques, and data sources to build future systems are already available today: the biggest challenge being how to build an appropriate bridge between them. In the following section, we consider a range of open problems that are hidden within this cycle of system building, and discuss how research is required to address them.

1.5 Looking Forward: Three Research Challenges

In the previous sections, we broadly characterised the what (data sources) and how (system models) of building systems with humans in the loop. As we move forward, it is likely that the applicability of these techniques will pervade many more facets of daily life; to date, many interesting datasets continue to remain behind closed doors. However, many open challenges remain; these range from technical to ethical challenges associated with computation with humans. In this section, we consider those open challenges that directly relate to building systems: a discussion of the broader issues of privacy and informed consent, while relevant, goes beyond the points that we enumerate here. Instead, we focus on three questions: (a) is the data itself valid? (b) does tailoring a system's design affect its inferences? and (c) how can the loop between data and user be closed most appropriately?

A prominent issue that arises when building the kind of (recommendation, retrieval, behaviour-related) systems discussed above is that understanding what is *not* represented in the data is often overlooked [14]. In this case, it is worth differentiating between data *sparsity*, which undermines the predictive power of machine learning algorithms, and data *representativity*, which is more about considering the inherent bias in the collection of *any* human dataset. Participation in publicly deployed applications, whether online or offline and with or without sensors, will be limited to self-selecting users; moreover, the typical distribution of participation between those who do tends to be highly non-uniform. Naturally, data derived from social media is 'skewed' by the extent that its users are a fair sample of the general population. While it is arguable that analysing data from millions trumps the same analysis on dozens, this newfound scale is not, in itself, a problem, but understanding the demographics of these sources is [40]. This issue gains importance if we consider that the way people use systems often breaks from our assumptions: notable examples include that online accounts (which, historically, researchers have assumed to belong to individuals) are shared between household members [5], and twitter data (which researchers assume is sourced from humans) is actually rife with automated bots [9].

From a system perspective, the interplay between the quantity and granularity of data collection continues to stand at odds with the obtrusiveness and energy consumption required on participants and their devices. For example, smartphone sensing applications will always need to trade off between sensor sampling rates and the battery usage; while recent work [47] shows how off-loading to nearby sensors may alleviate this problem, the question of how this design should vary between different contexts remains open. To what extent does optimising for battery life or, more generally, technical-related facets of a system (storage, connectivity, etc.) influence a system's ability to make reasonable inferences about a person's behaviour? Moreover, one of the most challenging tasks of system design is that of translating a high-level requirement (e.g., "I would like to give my users recommendations") into a task that can be suitably addressed by algorithms (i.e., "predicted preference ratings can be used to rank unrated content"). This process critically defines how systems reason about the data at hand and, more broadly, begins to put boundaries on the system's behaviour. In continuing with the recommendation example: defining an algorithmic approach to recommendations as one of generating a static set of predictions from user rating data will, for example, explicitly ignore the system's temporal behaviour [34].

Finally, the bridge itself between the data and systems continues to pose open research questions. Architecting a system to feed back any kind of information from human-sourced data requires researchers to continuously revisit the concept of relevance [54], not only in order to effectively close the loop, but also as a means of evaluating whether systems are worthwhile or achieving their goals [25]. This becomes particularly challenging in the context of the kind of potential systems described above: many of the methodologies for most appropriately evaluating the quality of the system (e.g., recommendation quality, whether/to what extent behaviour has been changed) remain elusive and the subject of active research.

1.6 Conclusion

In this chapter, we have provided a broad overview of how the *human sensor* may fit into the design of future systems. The main question that we discussed was, given the growing diversity and availability of data that encodes human behaviour: how can this data become an integral part of systems that support our every day life?

To do so, this chapter discussed the data itself, by considering the similarities in the different techniques that have historically been used to source it. Most notably, whether crowd-sourcing, participatory-sensing, or database mining techniques are adopted, the result is a representation of human behaviour which can be characterised in terms of its collection obtrusiveness (manual or automated), its original purpose (e.g., social vs. sensor), and its structure (both implicit hierarchy and whether behaviour is directly encoded or emerges from relationships in the data). We further broadly characterised three kinds of systems (recommendation, retrieval, behaviour-mediating) that can leverage these sources, and touched on three issues that remain unsolved within the context of system design. By attempting to thread a high-level narrative that demonstrates how human data can be put to use, this chapter naturally did not delve into the details of recommender, retrieval, and persuasive systems, although many suitable resources exist for further reading [4, 52, 19]. Finally, we discussed how these considerations motivate to future research: as future systems are designed, researchers and practitioners will need to tackle open questions about the input data itself, the potentially biasing role a system plays once high-level requirements are translated to algorithmic solutions, and how to most appropriately draw the link back to the humans that each system is designed for.

References

1. X. Amatriain. Mining Large Streams of User Data for Personalized Recommendations. *SIGKDD Explorations*, 14, December 2012.
2. L. Amini, E. Bouillet, F. Calabrese, L. Gasparini, and O. Verscheure. Challenges and Results in City-Scale Sensing. *IEEE Sensors*, 2011.
3. J. Ayers, B. Althouse, J. Allem, J. Rosenquist, and D. Ford. Seasonality in Seeking Mental Health Information on Google. *American Journal of Preventative Medicine*, 44(5), May 2013.
4. R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, 1999.
5. A. Bellogin, F. Diez, and I. Cantador. Time Feature Selection for Identifying Active Household Members. In *ACM CIKM*, Maui, Hawaii, Oct. 2012.
6. C. Brown, V. Nicosia, S. Scellato, A. Noulas, and C. Mascolo. Where Online Friends Meet: Social Communities in Location-based Networks. In *ICWSM*, Dublin, Ireland, 2012.
7. H. Bryan and P. Blythe. Understanding Behaviour Through Smartcard Data Analysis. *Proceedings of the Institution of Civil Engineers: Transport*, 160(4), 2007.
8. J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. Srivastava. Participatory Sensing. In *Workshop on World-Sensor-Web: Mobile Device Centric Sensory Networks and Applications*, Boulder, Colorado, 2006.

9. Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Who is Tweeting on Twitter: Human, Bot, or Cyborg? In *Proceedings of the 26th Annual Computer Security Applications Conference*, New Orleans, Louisiana, 2010.
10. S. Consolvo, D. McDonald, and J. Landay. Theory-Driven Design Strategies for Technologies that Support Behavior Change in Everyday Life. In *ACM CHI*, Boston, MA, April 2009.
11. S. Consolvo, D. McDonald, T. Toscos, M. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. Landay. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. In *ACM CHI*, Florence, Italy, 2008.
12. D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *WWW '09: Proceeding of the 18th international conference on World Wide Web*, 2009.
13. J. Cranshaw, R. Schwartz, J. Hong, and N. Sade. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In *ICWSM*, Dublin, Ireland, 2012.
14. d. boyd and K. Crawford. Six Provocations for Big Data. In *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011.
15. A. Das, M. Datar, A. Garg, and S. Rajaram. Google News Personalization: Scalable Online Collaborative Filtering. In *WWW*, Alberta, Canada, May 2007.
16. A. Doan, R. Ramakrishnan, and A. Halevy. Crowdsourcing Systems on the World Wide Web. *Communications of the ACM*, 54, April 2011.
17. N. Eagle and A. Pentland. Reality Mining: Sensing Complex Social Systems. *Personal and Ubiquitous Computing*, 10:255–268, 2006.
18. B. Ferris, K. Watkins, and A. Borning. OneBusAway: Results from Providing Real-Time Arrival Information for Public Transit. In *Proceedings of CHI*, Atlanta, GA, April 2010.
19. B. J. Fogg. Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity*, December 2002.
20. J. Froehlich, S. Consolvo, T. Dillahunt, B. Harrison, P. Klasnja, J. Mankoff, and J. Landay. UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits. In *ACM CHI*, Boston, USA, April 2009.
21. J. Froehlich and J. Krumm. Route Prediction From Trip Observations. In *Intelligent Vehicle Initiative, SAE World Congress*, Detroit, Michigan, 2008.
22. J. Froehlich, J. Neumann, and N. Oliver. Sensing and Predicting the Pulse of the City through Shared Bicycling. In *21st International Joint Conference on Artificial Intelligence*, Pasadena, California, 2009.
23. F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat. Digital footprinting: Uncovering tourists with user-generated content. *IEEE Pervasive Computing*, 2008.
24. P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh. WTF: The Who to Follow Service at Twitter. In *WWW*, Rio de Janeiro, Brazil, May 2013.
25. E. Hekler, P. Klasnja, J. Froehlich, and M. Buman. Mind the Theoretical Gap: Interpreting, Using, and Developing Behavioral Theory in HCI Research. In *ACM CHI*, Paris, France, April 2013.
26. T. Hossmann, C. Efstratiou, and C. Mascolo. Collecting Big Datasets of Human Activity One Checkin at a Time. In *Workshop on Hot Topics in Planet-Scale Measurement*, Lake District, UK, 2012.
27. S. Intille, J. Rondoni, C. Kukla, I. Ancona, and L. Bao. Context-Aware Experience Sampling. In *ACM CHI Extended Abstracts*, Ft. Lauderdale, Florida, April 2003.
28. V. Kalnikaite, Y. Rogers, J. Bird, N. Villar, K. Bachour, S. Payne, P. M. Todd, J. Schoning, A. Kruger, and S. Kreitmayer. How to Nudge In Situ: Designing Lambent Devices to Deliver Salience Information in Supermarkets. In *ACM Ubicomp*, Beijing, China, September 2011.
29. J. Krumm and A. Brush. Learning Time-Based Presence Probabilities. In *Pervasive*, San Francisco, California, 2011.
30. N. Lane, S. Eisenman, M. Musolesi, E. Miluzzo, and A. Campbell. Urban Sensing Systems: Opportunistic or Participatory? In *Workshop on Mobile Computing Systems and Applications (HotMobile)*, 2008.

31. N. Lathia and L. Capra. How Smart is Your Smartcard? Measuring Travel Behaviours, Perceptions, and Incentives. In *ACM International Conference on Ubiquitous Computing*, Beijing, China, September 2011.
32. N. Lathia and L. Capra. Mining Mobility Data to Minimise Travellers' Spending on Public Transport. In *ACM SIGKDD 2011 Conference on Knowledge Discovery and Data Mining*, San Diego, CA, USA, August 2011.
33. N. Lathia, J. Froehlich, and L. Capra. Mining Public Transport Usage for Personalised Intelligent Transport Systems. In *IEEE International Conference on Data Mining*, Sydney, Australia, 2010.
34. N. Lathia, S. Hailes, L. Capra, and X. Amatriain. Temporal Diversity in Recommender Systems. In *ACM SIGIR*, Geneva, Switzerland, 2010.
35. N. Lathia, D. Quercia, and J. Crowcroft. The Hidden Image of the City: Sensing Community Well-Being from Urban Mobility. In *Pervasive*, Newcastle, UK, 2012.
36. N. Lathia, C. Smith, J. Froehlich, and L. Capra. Individuals Among Commuters: Building Personalised Transport Information Services from Fare Collection Systems. *Elsevier Pervasive and Mobile Computing: Special Issue on Pervasive Urban Applications*, November 2012.
37. I. Li, J. Forlizzi, and A. Dey. Know Thyself: Monitoring and Reflecting on Facets of One's Life. In *ACM CHI Workshops*, Atlanta, Georgia, April 2010.
38. G. Linden, B. Smith, and J. York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEE Internet Computing*, 7:76–80, 2003.
39. G. Mackerron. Happiness and Environmental Quality. *PhD Thesis, The London School of Economics and Political Science*, 2012.
40. A. Mislove, S. Lehmann, Y. Ahn, J. Onnela, and J. Rosenquist. Understanding the Demographics of Twitter Users. In *AAAI ICWSM*, Barcelona, Spain, 2011.
41. A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo. A Tale of Many Cities: Universal Patterns in Human Mobility Modelling. *PLoS ONE*, 7(5), May 2012.
42. A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining User Mobility Features for Next Place Prediction in Location-based Services . In *IEEE ICDM*, Brussels, Belgium, December 2012.
43. D. Quercia, L. Capra, and J. Crowcroft. The Social World of Twitter: Topics, Geography, and Emotions. In *ICWSM*, Dublin, Ireland, 2012.
44. D. Quercia, J. Ellis, L. Capra, and J. Crowcroft. Tracking Gross Community Happiness From Tweets. In *AAAI CSCW*, Seattle, Washington, 2012.
45. D. Quercia, N. Lathia, F. Calabrese, G. D. Lorenzo, and J. Crowcroft. Recommending Social Events from Mobile Phone Location Data. In *IEEE ICDM*, Sydney, Australia, December 2010.
46. M. Ra, B. Liu, T. L. Porta, and R. Govindan. Medusa: A Programming Framework for Crowd-Sensing Applications. In *ACM MobiSys*, Lake District, UK, June 2012.
47. K. Rachuri, C. Efstratiou, I. Leontiadis, C. Mascolo, and P. Rentfrow. METIS: Exploring Mobile Phone Sensing Offloading for Efficiently Supporting Social Sensing Applications. In *IEEE PerCom*, San Diego, California, March 2013.
48. F. Radlinski and T. Joachims. Active Exploration for Learning Rankings from Clickthrough Data. In *In proceedings of KDD*, 2007.
49. C. Ratti, R. M. Pulselli, S. Williams, and D. Frenchman. Mobile Landscapes: Using Location Data From Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design*, 33(5):727–748, 2006.
50. C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. Strogatz. Redrawing the Map of Great Britain from a Network of Human Interactions. *PLoS ONE*, 5, December 2010.
51. S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using Mobile Phones to Determine Transportation Modes. *ACM Transactions on Sensor Networks*, 6(13), February 2010.
52. F. Ricci, L. Rokach, B. Shapira, and P. Kantor, editors. *Recommender System Handbook*. Springer, 2010.

53. C. Roth, S. Kang, M. Batty, and M. Barthelemy. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLOS ONE*, 6, 2011.
54. T. Saracevic. Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science*, 26(6):321–343, 1975.
55. B. Sarwar, G. Karypis, J. Konstan, and J. Reidl. Item-Based Collaborative Filtering Recommendation Algorithms. In *WWW*, Hong Kong, China, 2001.
56. B. Shaw, J. Shea, S. Sinha, and A. Hogue. Learning to Rank for SpatioTemporal Search. In *Web Search and Data Mining (WSDM)*, Rome, Italy, 2013.
57. V. Soto and E. Frias-Martinez. Robust Land Use Characterization of Urban Landscapes using Cell Phone Data. In *Workshop on Pervasive and Urban Applications*, San Francisco, CA, June 2011.
58. Y. Takeuchi and M. Sugimoto. An Outdoor Recommendation System Based on User Location History. In *ACM Ubicomp*, 2006.
59. L. Weinstein. TfL's Contactless Ticketing: Oyster and Beyond. In *Transport for London*, London, UK, September 2009.
60. R. White, N. Tatonetti, N. Shah, R. Altman, and E. Horvitz. Web-Scale Pharmacovigilance: Listening to Signals from the Crowd. *American Journal of Preventative Medicine*, March 2013.
61. T. Yan, M. Marzilli, R. Holmes, D. Ganesan, and M. Corner. mCrowd: A Platform for Mobile Crowdsourcing. In *ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Berkeley, California, 2009.
62. X. Yu, Q. Fu, L. Zhang, W. Zhang, V. Li, and L. Guibas. CabSense: Creating High-Resolution Urban Pollution Maps with Taxi Fleets. In *ACM MobiSys*, Taipei, Taiwan, June 2013.
63. J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun. Where to Find My Next Passenger? In *ACM Ubicomp*, Beijing, China, September 2011.
64. P. Zhou, Y. Zheng, and M. Li. How Long to Wait? Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing. In *ACM MobiSys*, Lake District, UK, June 2012.