Putting the feedback cycle in high gear: community-sourced, data-driven approaches for sustainable transportation infrastructure

Randy H.Katz, UC Berkeley and Lise Getoor, UC Santa Cruz

The challenge of modifying physical infrastructure and policies to meet societal goals would benefit from a virtuous feedback loop in which policy makers are able to use data collected from communities to choose the most impactful designs, and the resulting changes motivate the community to provide ongoing, long-term data. We study this challenge in the context of the concrete model problem of changing travel modes to reduce the impact of transportation - the largest source of greenhouse gas (GHG) emissions in California. Prior efforts at mode shift have focused on single modes of transportation, incentives for changing personal behavior, or aggregate utility model generation. We argue that a more holistic view that combines low power, passive data collection, advanced user and community modelling, and qualitative data gathered through active learning has the potential to generate such a feedback loop, but introduces a host of new challenges in both the systems and modelling areas. Our collaboration between the systems-focused UCB team with an existing system for data collection and analysis, and the modelling-focused UCSC team with an existing scalable statistical modelling toolkit is well equipped to handle these challenges. We plan to deploy our combined solution in a series of pilot projects on the UCB and UCSC campuses, thus exploring the challenges with mode shift in two very different settings. We can use these pilots as the basis for submissions to solicitations associated with the new Presidential Initiative on Smart Cities.

Introduction

Improvements to societal infrastructure, which are vital to achieving sustainability, typically proceed through a long, mostly uninformed process with little real feedback from the target communities; consequently, all too often these improvements underperform or are poorly received. The pervasive use of Information and Communication Technology (ICT) for sensing and the use of advanced machine learning techniques can make an entirely different development process possible. We propose the creation of an information nexus that can lead to the formation of naturally connected communities around potential infrastructure improvements. We can sense existing usage patterns from such groups, understand the choices underlying these patterns by querying individuals directly, solicit feedback on proposed changes and observe the effect of deploying prototypes on actual usage patterns. This proposal builds upon existing technology developed by UCB for multimodal mobility data collection and analysis using smartphones, and by UCSC for machine learning-based socio-behavioral modeling. Bringing these two lines of work together allows us to tackle a concrete 'model problem' of supporting the use of *Active Transportation* (e.g. walking and biking) whose health, economic and quality of life benefits are well established. By doing so, we will provide both a technical and developmental proof point that is central to UC's zero emission goal and California's larger climate goals.

Societal Change Model Problem

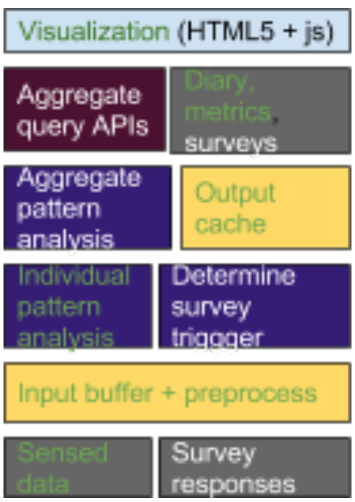
UCB's Dept. of Parking and Planning approached us to see if our existing work on smartphone-based multi-modal data collection and analysis could be utilized to help increase the bicycling mode share to campus. This will help UC make progress on its carbon neutrality goals, while reducing the space diverted to parking in a dense urban campus. UCB and UCSC present two very distinct active transportation challenges, one being an urban campus immersed in numerous challenging byways and the other being connected primarily through public transport. The goal is to determine which investments - bike lanes, secure bike parking, e-bike discounts, etc. - are likely to have the greatest impact on each campus. The fundamental challenge is that in the absence of accurate information about what people do now, what are the barriers to adoption, and how people react when the barriers are removed, these investment decisions have to be made on a speculative basis instead of being user and data-driven.

We plan to develop and deploy several interconnected components to help answer these questions. A *bpedal* phone app, web resources, and analytics infrastructure will be developed and promoted as a means of connecting the community to the campus effort. It will allow users to glean and share current behavior by collecting travel patterns across all modes of transportation (e.g. motorcycle, scooter, electric bike). This data will provide personalized feedback to users on metrics such as their carbon footprint and daily physical activity. To identify barriers to adoption, we will augment the sensed data with qualitative data on safety and scripted survey engagements on proposed improvements. Sharing our aggregated data with the campus community will allow us to identify problem areas, observe changes in travel patterns resulting from improvements, and provide a conduit for suggesting and prototyping improvements.

Technical Challenges

The world of smartphone tracking apps is vast, including fitness (Google Fit, Apple Health), route sharing (RunKeeper, Strava) and trip planning (Google Maps, City Mapper). Efforts to motivate users to reduce emissions exist in limited government initiatives (e.g., EU's peacox, matkahupi). In contrast to these efforts, we offer two distinct advantages: impact and personalization. Working with the planning department allows us to translate user experience and feedback into policy decisions. Our analytical approach will also provide a personalized experience for individuals and identify key communities. Together, these aims enable us to expose the virtuous feedback loop of personal activity awareness, community sharing, and engaged planning.

Seamless integration across these levels presents new technical challenges. Unlike traditional transportation surveys, we must provide value to users so that they are engaged and continue to provide high quality data over extended periods. User behavior modelling and community identification can help us identify value to the user. Unlike personal tracking apps that "are unbothered by … standards of evidence", data that is used for policy decisions needs to come with estimates on accuracy. Machine learning approaches can model mobility patterns to combat noisy measurements from mobile devices. The ground truth required for accuracy evaluation must ultimately come from the user, but user attention is arguably the most precious resource today. Active learning, a machine learning problem that identifies the most useful information to acquire, is an open area of research in this domain.



These and other challenges have led us to develop the system architecture shown here. While some components are represented in existing efforts, this setting presents unique challenges. Data collection must not place undue burden on individuals; thus, sensing should be automatic and extremely low power, yet accurate. We have carefully analyzed the state-of-the-art, found that it is inadequate, and developed novel uses of existing APIs to move it forward. Individual streams must be unaffected by intermittent connectivity, many streams must be multiplexed in the analytics infrastructure, and key data should be propagated back to the phone to allow localized use of model parameters. Survey questions should be updateable by planners as they evaluate changes, and the surveys should be triggered in a way that is most likely to generate high quality responses from the users, i.e., be driven by analysis of the data. Users must derive value from the collected data; this involves both distillation into a personalized summary and actionable insights and useful derivatives of community-wide data and institutional plans. The structural level involves aggregating individual data and analyze overall patterns to generate similar summaries and insights, but also model formation of individuals both to offer better recommendations and to enable simulation of proposed alternatives.

A common thread in these challenges is the ability to build modules at different scales. User-level models need to capture longitudinal patterns in an individual's mobility traces and the feedback most likely to maintain the individual's engagement over time. Group models need to aggregate data across many individuals to find meaningful patterns in transportation usage. These patterns include frequent routes, common bottlenecks or safety issues, and correlations between points of interest and transportation choices. Together the user and group models allow us to tailor the experiences of users as well as optimize surveying and ground truth acquisition to the most informative instances.

Technical Readiness

Building social and technical solutions to these challenges requires leveraging the key strengths of both the UCB and UCSC teams, and their ongoing projects.

The systems-focused UCB team has built a series of prototypes for mobility monitoring with support for active transportation modes, tour model generation, and model-based recommendation. This open source system consists of low power, automatically triggered data collection on both iOS and Android, and a server-side pipeline for cleaning, segmentation and transportation mode inference. The resulting daily timeline and daily summary metrics are displayed on the user's smartphone with buffering to handle slow or disconnected operation.

The modeling-focused UCSC team has developed probabilistic soft logic (PSL), a highly-scalable, open source statistical modeling toolkit that has been applied to a variety of information integration and socio-behavioral prediction problems. PSL's key strength is statistical relational learning: the ability to leverage the relationships between variables and produce more accurate predictions. PSL has been used to identify latent groups and their concerns, such as student complaints in online courses (Ramesh et al., ACL15), and infer useful facts from noisy Web extractions (Pujara et al., AIMag15).

Our proposed solution will combine the UCB team's deep experience in collecting and understanding mobility data using mature systems infrastructure with the UCSC team's tradition of building sophisticated models that achieve state-of-the-art results.

Timeline and Additional Funding

We plan to implement the project in three phases. In **Spring 2016**, we will deploy a pilot study of < 100 individuals at UC Berkeley that will collect both passively sensed and limited survey data. The focus will be on active transportation modes, and we will make the resulting heatmaps available to both the planning department and to the community at large. In **Summer 2016**, we will extend the data collection to cover a rich variety of modes, implement more complex surveys by integrating with ohmage-omh, and explore expanding the user base to UCSC. In parallel, we will start building user models based on data collected in Spring. In **Fall 2016**, we will integrate user models into community models and use combined models for active learning and recommendations.

Drawing upon previously submitted proposals to Arpa-E and NSF, we will be utilizing this new study as the basis for submissions to solicitations associated with the new Presidential Initiative on Smart Cities, which include $2.5 million for the Global City Teams Challenge and $10 million for NSF CPS research awards focused on Smart and Connected Communities.