Demo Abstract: Extrapolation from Participatory Sensing Data

H. Liu[1], S. Gu[1], C. Pan[1], W. Zheng[2], S. Li[1], S. Hu[1], S. Wang[1], D. Wang[1], T. Amin[1], L. Su[1], Z. Xie[3], R. Govindan[4], A. Barnoy[5], T. Abdelzaher[1]

[1]University of Illinois at Urbana-Champaign, [2]University of Wisconsin-Madison, [3]University of Virginia
[4]University of South Califonia, [5]City University of New York

ABSTRACT

In this demo, a learning system, called *Metis*, is presented that extrapolates missing pieces in participatory sensing data. The work addresses the challenge of *incomplete coverage* in participatory sensing applications, where lack of complete control over participant mobility and sensing patterns may create coverage gaps in space and in time. Metis learns the underlying spatiotemporal patterns of the measured phenomenon from available incomplete observations, and uses these patterns to infer missing data. We describe the overall system design and demonstrate the system using data collected during the New York City gas crisis in the aftermath of Hurricane Sandy.

1. INTRODUCTION

We consider participatory sensing applications, where the state of several spatially distributed points of interest (PoIs) must be monitored over time. We assume that the state can be represented by one or more bits per PoI. Unfortunately, our participants offer only sporadic coverage. Hence, the state of some PoIs may be unknown at any given time. An important challenge for the participatory sensing system is therefore to fill-in the missing data.

An example of such a system may be the disaster response application shown in Figure 1, where volunteers and first responders must survey the damage in the aftermath of a mass-destruction event. In this demo, we shall focus on the aftermath of hurricane Sandy, where many gas stations around New York City lost power and/or gas. An important question became to map out those stations that remained operational. Unfortunately, due to loss of power and communication, mapping operational gas stations was challenging. Information on gas availability was not always available everywhere.

Fortunately, the state of PoIs in the physical world is often correlated. For example, failures of some gas stations were

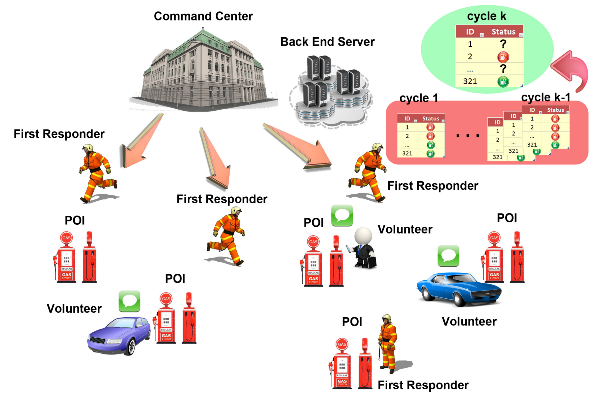

Figure 1: Application scenario.

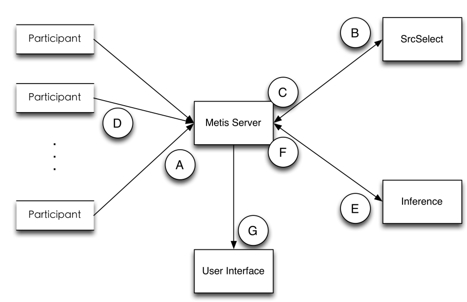

Figure 2: *Metis* System Overview.

correlated in the aftermath of Sandy. Gas stations that were on the same power line tended to run out of power together. Similarly, gas stations that used the same suppliers tended to run out of gas in a correlated fashion. These correlations among physical world states can be learned over time in order to offer predictions that fill in coverage gaps.

Figure 2 presents an overview of our proposed system that aims to extrapolate from available observations in order to fill-in coverage gaps. The system collects the status of PoI sites (e.g., in a disaster region). Participants report PoI state to the back-end server (event A). The server then continuously infers correlations among PoI sites based on collected data.

When information of some PoI sites is missing at a given

SenSys '13, Nov 11-15 2013, Roma, Italy.

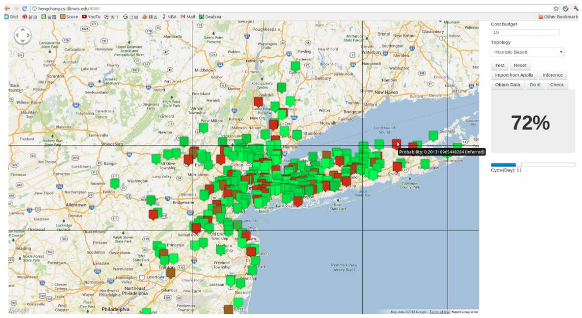

Figure 3: User interface.

time, *Metis* infers that information by calculating a probability distribution of the state of each site. Two components, *Inference* and *SrcSelect*, play an important role in this process.

The *Inference* component takes the current partial state of PoI sites, as well as historical data, as inputs (event E), and infers the state of other PoI sites in a probabilistic manner together with confidence levels. After that, the inferred results are sent back to the *Metis* server (event F) and the server provides a complete PoI map to system users (event G). We use a Monte Carlo Maximum Likelihood Estimation-based inference approach.

The *SrcSelect* component runs when the resources available (e.g., first responders) to collect observations have a limited budget and are unable to cover all PoI sites. It takes into account the benefit of obtaining information from each of the PoI sites, the cost for collecting information from each PoI, and the budget that the service provider currently has (event B). Based on the above, it calculates the most informative subset of PoI sites (event C), in that investing in finding out their actual state would maximally help infer the state of other nodes in the network. Those PoIs would then be the target of focused data collection. An advantage of Metis is that it finds the minimal subset of PoIs to query, such that the cost of data collection is minimized.

2. DEMONSTRATION

We fully implemented *Metis* and the implementation is completely modular. Our system is divided into four major components: sensory data collection, inference engine, source selection module, and user interface.

Figure 3 shows the user interface when applying *Metis* to the gas station shortage application after Hurricane Sandy struck NYC. Information on each POI (i.e., gas station), either collected or inferred, is shown to system users in real time. Color coding is used to differentiate the probabilities of inferred results (e.g., probability of gas availability).

The demonstration starts with the *NYC Mission*, where we have ground truth on gas availability from two data traces. The first data trace was collected by the All Hazard Consortium (AHC) [2]. It covered gas stations in all cities that were affected by Sandy, in states including WV, VA, PA, NY, NJ, MD, and DC. The information was updated daily. We recorded the status of these gas stations over 32 days, from November 2nd to December 2nd, and found that there were 321 stations that were unavailable for at least one day. We chose these 321 gas stations as the POI sites in our *Metis* system, and their status on 32 days as the input data. The locations of these gas stations are displayed in Figure 3. The second data trace came from GasBuddy [3]. During the Sandy disaster recovery period, GasBuddy established a participatory sensing service to collect gas station information in the disaster region. This data trace contains information on 254 gas stations collected by volunteers in an opportunistic manner over 21 days, and 56 of them are overlapped with the AHC trace.

Given the above two traces, users on the NYC mission can select any date from within the 32-day period, and indicate the number of gas stations whose ground truth status is to be "revealed" to *Metis*. This sets up the current time as well as the currently available information for the mission. The user assumes the role of a mission commander. The mission is to determine the status of the remaining gas stations.

In the demo, users will be allowed to enter a "competition" with Metis, by choosing either the *manual* or *assisted* demo mode. In the manual mode, the user can view the past known (and incomplete) history of gas availability up to the current date and manually decide which gas stations have gas today. The user will also be allowed to send a "patrol" to reveal ground truth about selected gas stations (e.g., those that they have trouble guessing). Using that information, the user may guess as much missing state as they can. When done, the user is scored by the (i) accuracy of guessing, (ii) the number of remaining unknowns, and (iii) the number of patrols used.

In the assisted mode, *Metis* is activated to infer missing state automatically by clicking the *Inference* button. Our algorithm infers the status of additional gas stations subject to a confidence threshold based on the same information furnished in the manual mode. If the resulting coverage does not satisfy the user, the user can "send patrols" (as before) to reveal the actual status of more gas stations. The optimal selection of gas stations to send the patrols to is computed by hitting the *SrcSelect* button. It identifies the most beneficial subset of gas stations to query that gives the most insight into the values of missing data. With additional ground truth uncovered, inference can be run again. When the user is satisifed, they can ask that the result be scored (for accuracy, missing information, and number of patrols used) by clicking the *Check* button. Users will be challenged to beat the performance of the assisted mode.

Other missions will be made available (besides the NYC mission) based on synthetic data on fake disasters. Users will also be allowed to change the algorithms exploited for inference and for source selection in the assisted mode, to get insights into their relative merits and weakenesses in different missions.

3. REFERENCES

[1] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In World Sensor Web Workshop, ACM Sensys 2006, Boulder, Colorado, October 31, 2006.

[2] All Hazards Consortium. URL: `www.ahcusa.org/`.

[3] GasBuddy: Find Low Gas Prices in the USA and Canada. URL: `www.gasbuddy.com/`.