PEOPLE FORECASTING: RESEARCH PROJECT DESCRIPTION

Jonathan Hutchins

1. PROJECT MOTIVATION AND PROBLEM STATEMENT

My thesis work began as part of project Rescue (http://www.itr-rescue.org) whose goal is to help emergency personnel better respond to natural and man-made disasters. I titled my Rescue subproject ***"People Forecasting"***. Just as a weather forecast helps people make more informed decisions about what to wear, where to go and when to go there; the goal of People Forecasting is that emergency personnel would use our models to discover information about population density and movement city-wide at any instant in time, resulting in increased speed and improved quality of their decisions regarding response efforts during a disaster.

Human activity sensors, such as pictured in Figure 1, are becoming more common and are forming an increasingly dense network, especially in urban areas. They provide real-time information about the presence and movement of people that could be valuable in emergency response, transportation planning, and security applications. The aim of my research project is to extract useful information about human activity and behavior that is hidden in these simple measurements.

This project is particularly appealing to me because I believe our research has potential to be of great value to the emergency response community. The existing state-of-the-art methodology for occupancy estimation on a large-scale; for example, relies on census information (e.g, in transportation and urban planning studies), sometimes with a constant multiplier for daytime for industrial and residential areas. However to our knowledge, there has been no systematic approach to using information from human activity sensors to provide real-time, qualitative information about population behavior which could be of use to first responders. Census information can't be used to predict whether Dodger's stadium is full or empty, but traffic sensors nearby might give an indication.

Fig. 1. The left panel shows an optical people counter that we installed at the CalIT2 building on UCI's campus. Loop detectors maintained by the California Department of Transportation on Southern California freeways are shown in the right panel (measurements obtained via PeMS). Sensors such as these report measurements of human activity in real-time.

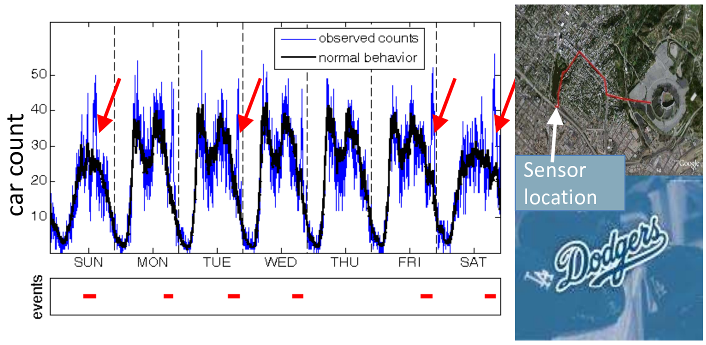

Fig. 2. One week of traffic measurements (blue line), with the normal traffic pattern (black line) learned by our model. Unusual event activity such as indicated with red arrows correspond to home games of the LA Dodgers Baseball team (game times indicated in bottom panel. The presence of unusual event activity corrupts the normal pattern causing difficulties for methods attempting to learn this underlying normal patterns.

Our work has provided a completely new (and potentially much more accurate and fine-resolution) approach to problems such as occupancy estimation; making use of dynamic information such as is provided by people count sensors and loop detectors.

2. LEARNING PATTERNS OF HUMAN BEHAVIOR AND PREDICTING UNUSUAL EVENTS

Sensors that record counts of observations of a human activity have the advantage of being relatively low-cost and privacy-preserving as compared to video, and are also commonly found in large networks city-wide. The data (see Figure 2) contains a variety of different phenomena including (a) systematic hourly, daily, weekly, and seasonal effects, (b) evidence of a wide variety of events and unusual counts (such as caused by large sporting events or concerts), and (c) various types of sensor noise and faults (such as over-counting and sensor failures). The volume of the data available, and the fact that the data are being recorded in a real-world environment, provides very useful challenges for testing a variety of theories and models. As such this large sensor data set is ideal for experimentation and algorithm development.

Learning the predictable patterns of human behavior and detecting and quantifying unusual activity could have great value to emergency planners and responders. Information about predictable human behavior can be used to help fill in missing measurements, predict future behavior, or aid in emergency planning such as evacuation planning. Information about unusual activity is particularly useful in situation assessment and resource allocation. However, the bursts of unusually high activity such as pictured in Figure 2 and unusually low activity caused by traffic accidents for example; corrupt the underlying pattern and are prevalent enough to make the problem (of learning the underlying behavior and prediction unusual activity) non-trivial. This problem is fundamentally challenging because the sensor data is typically unlabeled, i.e., the available training data is not separated into normal behavior and abnormal events - to find the abnormal events requires knowledge of normal behavior, and to learn normal behavior requires that one can first remove the abnormal events. Our contribution has been to solve this "chicken-and-egg" problem (for time-series counts of human behavior) using a robust, general, and systematic statistical framework.

In our early work [3, 4] we presented a general statistical framework for unsupervised learning of both recurrent patterns of human behavior over time as well as characteristics of unusual events found in time-series of sensor count measurements. In addition, we have tested and validated the methodology on several months worth of (a) people-counter data from the UCI Calit2 building, and (b) loop-sensor traffic data from Los Angeles freeways. Our method was shown to be significantly more accurate ad detecting unusual activity than simple threshold models. Our model also infers other important attributes of unusual events that might be of interest to first responders including the duration of the event and the number of extra cars/people attributed to the unusual event. It is not clear how these attributes could be estimated using

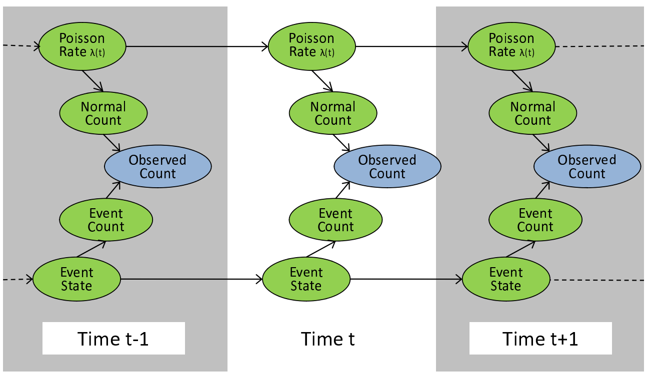

Fig. 3. Graphical model of our general framework for modeling human activity sensors. Both the event and rate variables couple the model across time: the Markov event process captures rare, persistent events, while the Poisson rate parameters capture the underlying normal activity.

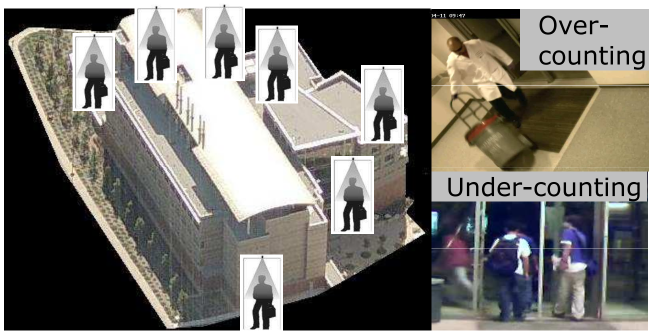

Fig. 4. The doors to the CalIT2 building on UCI's campus were instrumented with optical people counters; and the probabilistic framework described in Section2 was extended to infer building occupancy. The new model included the concept of measurement noise as illustrated in the panels to the right showing examples of over-counting (top) and under-counting (bottom). The output of the building occupancy model was shown to be robust, even in cases of missing and corrupted measurements.

simpler approaches.

Our approach is based on a combination of hidden Markov models and Poisson processes (the graphical model representation of our method is seen in Figure 3). This work has been well-received in the data mining and sensor data analysis research communities - for example our 2006 ACM SIGKDD conference paper[3] was one of only 8 papers from the conference selected for subsequent journal publication in the ACM TKDD Transactions[4], out of approximately 400 papers that were originally submitted to the conference.

3. MODELING BUILDING OCCUPANCY

We extended the general framework described in Section 2 to develop new techniques for estimating how many people are in a building as a function of time given noisy entrance and exit counts from door sensors. To evaluate our method, we instrumented the doors to the CalIT2 building on UCI's campus with optical people counters (see Figure 4).

Simple counting techniques tend to become severely biased over time due to sensor noise (Figure 4). In [1] (Hutchins, Ihler, Smyth, 2007) we showed that our proposed approach provides much more accurate estimates than the simpler baselines, particularly in cases where sensor measurements are not reported due to communication failures for example, and in cases where the measurements are corrupted.

In addition to adding the concept of sensor noise to the model, the building occupancy model showed that individual sensor models of the type described in Section 2, could be linked together in a meaningful way to learn additional information that could not be learned from the models independently. The algorithms and tools we have been developing can be used as components of much larger situation assessment tools that can quickly and accurately assess how many people are in a given location from very noisy sensor measurements and also determine whether any unusual patterns are present or not.

4. LARGE SCALE URBAN ANALYSIS

Large-scale loop sensor data such as found in our urban case study are well known to transportation researchers, but have resisted systematic analysis due to the significant challenges of dealing with noisy real-world sensor data at this scale. Our approach removes faulty measurements and discovers the underlying signal in a much more general way than is currently practiced.

We performed a large-scale analysis of an urban traffic sensor data set (see Figure 5). Because of the underlying probabilistic framework of our approach, it is relatively straightforward to extend and scale our methodology to new problems and new sensors. We applied our model to over 1700 loop detectors (with 100 million sensor measurements over several months) in Los Angeles and Orange County, CA [2] (Hutchins, Ihler, Smyth, 2008). This case study presented our model with a large variation of measurements, including sensors with extended periods of missing measurements (no measurement reported) and sensor failure.

There were a number of cases where our original model was not effective at finding the underlying pattern of human behavior. A fault-tolerant extension of the original model addressed this shortcoming by adding an additional Markov chain whose state variable indicated the presence of a sensor failure.

This case study allowed us to make city-level queries of our data and also to perform spatial analysis. The model was able to provide useful insights about a data set that is widely considered difficult to analyze (e.g., see the recent 2007 Statistical Science paper from the Berkeley Traffic group).

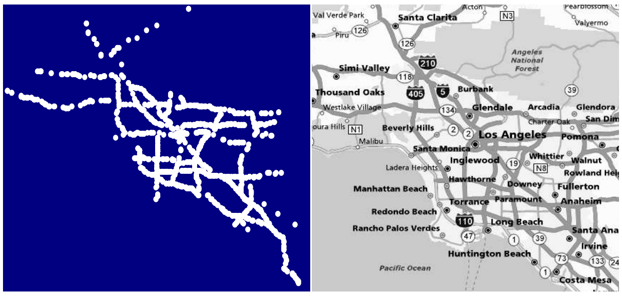

Fig. 5. The left panel shows the sensor locations of the 1716 freeway loop detectors that were modeled in the urban scale-up study. The sensors covered Los Angeles County and Orange County as seen in the map on the right.

5. FUTURE WORK

I am currently integrating our research into a software system called SATware that is a product of project Rescue. Real-time implementations of the models described in Sections 3 and 4 will be incorporated into this environment, and will allow the first-responder community to evaluate our research.

This fall I will begin a joint work with the Institute of Transportation Studies (ITS) at UCI. I presented the work described in [2] (Hutchins, Ihler, Smyth, 2008) to several members of the ITS group who work with the same group of sensors we modeled in Section 4. They are interested in integrating our framework into some of their work focusing on detecting the temporal and spatial signatures of traffic accidents. This application will require another extension of our model in order to link the sensors spatially.

We would also like to create a denser network of sensors that detect human presence in the CalIt2 building and other buildings on campus which could be used in the development of tracking algorithms. Such a large network of sensors across a university campus would allow us to extend our models further to learn and predict additional patterns of movement over a large area.

Another interesting direction is large-scale dynamic population density estimation. This was one of our original goals for the traffic data set. The fault-tolerant extension of the model described in Section 4, has helped us overcome some of the initial roadblocks we encountered with this important problem, especially in the context of situation assessment and emergency response.

This research summary presents a general statistical framework for learning valuable information that is hidden in human activity sensor measurements; and shows its flexibility to be extended to new research problems and applications as illustrated in Sections 3 and 4. As human activity sensors become more dense, other research opportunities and applications will emerge in emergency response, transportation planning, and security that are not currently possible; and probabilistic frameworks such as ours will provide tools for addressing these new challenges.

6. REFERENCES

[1] J. Hutchins, A. Ihler, and P. Smyth. Modeling Count Data from Multiple Sensors: A Building Occupancy Model. *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pages 241–244, 2007.

[2] J. Hutchins, A. Ihler, and P. Smyth. Probabilistic analysis of a large-scale urban traffic data set. *Proc. of the 2nd International Workshop on Knowledge Discovery from Sensor Data (ACM SIGKDD Conference, KDD-08)*, 2008.

[3] A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying Poisson processes. In *ACM Int'l Conf. Knowledge Discovery and Data mining*, pages 207–216, 2006.

[4] A. Ihler, J. Hutchins, and P. Smyth. Learning to detect events with Markov-modulated poisson processes. *TKDD*, 1(3), 2007.