Global Priors of Place and Activity Tags

Donald J. Patterson
Department of Informatics
University of California, Irvine
djp3@ics.uci.edu

Abstract

This paper describes an approach for creating detailed full-coverage labellings of human activity. Our goal is to create global maps of physical positions labelled with a distribution over the most likely place name and most likely activity. We ground our ontology of labels as: the term that a person would want to display to someone before they initiate a communication. Rather than compiling a canonical list of possible labels, we piggyback the label data collection in a situated communicative exchange. Using ideas inspired by image segmentation and extended to support our goals we propose machine learning techniques for smoothing distributions across gaps in existing data.

Introduction and Related Work

The consumerization of sensor-laden platforms such as mobile phones, laptops, and vehicles, that also provide access to the collected data, is enabling data sharing and activity reasoning to scale to new levels. A difficult challenge which remains is to understand how the sensors from different users, ostensibly deployed for specific different reasons, can be principally combined and aggregated for new types of uses.

In this paper we introduce a data collection approach and inference mechanism by which it may be possible to create dense geographic maps that are identified with place and activity labels. We obtain information about place and activity labelling in a supervised manner by piggybacking onto the communication practices of instant messaging and cellphone users.

Related work on plan recognition (Kautz and Allen 1986; Wilensky 1983; Colbry, Peintner, and Pollack 2002), vision-based activity recognition (Jebara and Pentland 1999; Boger et al. 2006; Shi et al. 2004), and object-interaction-based activity recognition (Perkowitz et al. 2004; Philipose et al. 2004) formulates activities as collections of stereotypical sequential actions. When these approaches are applied to real-world data, they tend to be successful in modelling low-level behaviors in controlled environments or in very domain-specific applications (*e.g.*, hand-washing). We believe that when applied to global scale environments these approaches will be of limited success. The wide variety of valid approaches to achieving a goal (executing an activity) and correspondingly, the wide variety of reasons why such an activity might want to be recognized, make a global state-based recognition system a daunting pursuit. For example, successfully recognizing that someone has “made tea” has important modelling implications if the application automatically calls social partners to the table as opposed to ordering new tea bags or monitoring caffeine intake.

An alternative approach in the literature has been to treat activity recognition as a classification problem with less strict interpretation, if any at all, of the steps involved. This includes unimodal evaluations of activity and social context from audio (Choudhury and Pentland 2003; Stäger et al. 2003; Stäger, Lukowicz, and Tröster 2004), video (Fitzpatrick and Kemp 2003), accelerometers (K. Van Laerhoven and Gellersen 2004) and RFID (Patterson et al. 2005).

Although in this work we also approach the activity recognition task as one of classification, we emphasize multi-modal sensor evaluations like (Kern et al. 2004) and (Choudhury, Lester, and Borriello 2005) in order to avoid biasing our effectiveness on activities that are easily discriminated by one sensor (*e.g.*, choosing activities such as “hammering” or “grinding coffee” because sensing is done with a microphone).

In this paper we describe a system for collecting place and activity data from users. We characterize the existing data and propose two models for smoothing the estimations of place and activity names using loopy belief propagation. Additionally we suggest some relevant extensions that are necessary in order to account for application specific effects. Our goal is to use our data to create a dense map of priors over place and activity labels for every location on earth.

System

In determining the appropriate *place* or *activity* label for a geographic *position* it is important to ground the labelling in a specific practice, both to understand the nature of the data collected and the purposes to which it can reasonably be put (Hightower 2003). For the work described in this paper, we developed an instant messaging (IM) application and a Blackberry mobile phone application both of which function like a smart address book. These applications provide access to a buddy list and a contact list, respectively, in which each

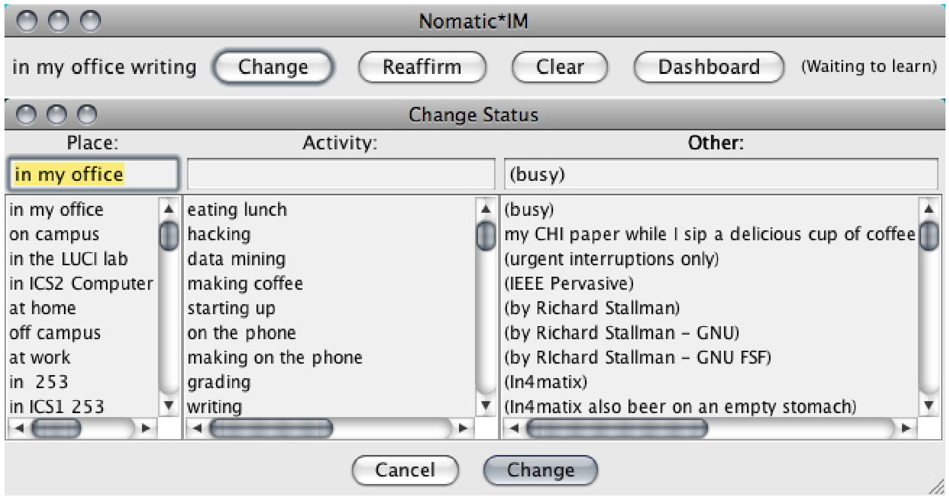

Figure 1: **top**: Nomatic user interface before the user hits "Change" **bottom**: Subsequent suggestions provided by Nomatic tool in lieu of current sensor readings.

name is augmented with a textual description of what the remote buddy/contact is currently doing.

In order to be able to have such information available the remote contact must manually update their textual description on their computer or phone. We have made a software tool called Nomatic (Patterson, Ding, and Noack 2006) which makes the process more automatic by acting like a context-aware experience sampling system (Intille, Kukla, and Ma 2002). Nomatic watches sensors on the device and when it detects a significant change in context it prompts the user to describe their current place, and activity. In addition to simply prompting a user, Nomatic also uses user-specific decision trees to suggest appropriate labels for quick entry. Instead of benefitting a researcher, however, this label is broadcast to contact lists to help mitigate inappropriate interruptions (see Figures 2 and 1).

In the process of updating their place and activity status, this information, paired with sensor readings, is sent to a community history database which forms the data on which the work in this paper is based.

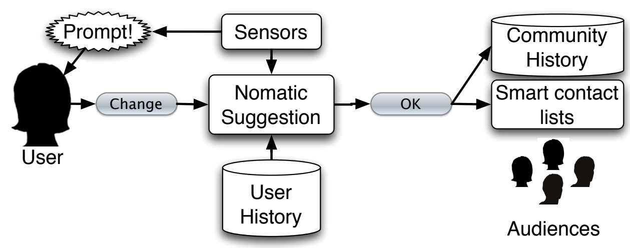

Figure 2: **Control Flow of Nomatic Tool**: A user wants to update their status message on their contact list, possibly due to a context-aware prompt, so they hit a "Change" button. Nomatic scans current sensors and uses machine learning techniques to suggest a list of statuses from their personal history. After a user picks one it is sent to IM buddy lists and contact lists for viewing by remote contacts. The data is also stored in a community history database for the work described in this paper.

Data

We currently have over 40 laptop users and 4 Blackberry users who are contributing data to our community history. The data has been collected over 13 months and contains over 600,000 status entries paired with sensor readings, predominantly from laptops.

In Table 1, we show the most likely, median likely, and least likely status entries. This table demonstrates the wide variety of ways in which people choose to categorize their place and activity. The activities are not exclusive (*e.g.*writing vs. working) do not look like the activity classes that are currently represented in the academic literature and include a range of detail, granularity, and specificity.

Figure 3 shows a histogram of the most popular place and activity entries. Of the 487 unique place labels, the top 10

Rank	**Place Label**	**Activity Label**
1	at home	{blank}
2	{blank}	writing
3	in my office	working
4	at dorm room	writing a paper
N/2	my friend's house	drinking green tea
N-3	Brixton	looking for a place to live
N-2	Shanghai	solving problems
N-1	Olympia, Washington	making flyers
N	Lugu Lake	sight seeing

Table 1: A table with typical status label entries. The place and activity labels are independent.

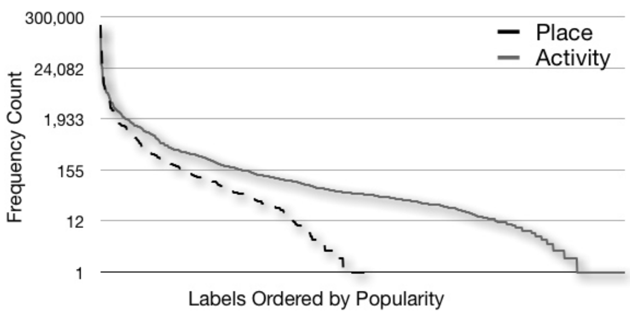

Figure 3: A histogram of label popularity. Each element on the x axis corresponds to one unique place or activity label ordered from most popular on the left to least popular on the right. In this graph place and activity are independent.

account for 72% of all place entries. Of the 947 unique labels, the top 10 account for 51% of all activity entries. This data demonstrates the long-tailed nature of the ways that people use to enter their status information.

Given this wide variety of data, it is possible to consider generative approaches which treat activity as a language construction process that is guided by current sensors.

Modelling

From this data we wish to develop priors of place and activity labels. We approach this task by modelling our data in two ways. One as a GPS grounded data set and the second as a network of connected wifi access points. Both of these techniques are necessary because not all of the devices that are contributing data to our database are outfitted with GPS devices or conversely with wifi APS.

GPS Modelling

For GPS Modelling, we treat the world as a grid of 10m blocks. We assume that each grid cell has a hidden "most appropriate" place label and independent "most likely" activity label. Clearly there are some obvious faults with this assumption, such as the cell boundaries not being correctly sized and aligned with the hypothesized true boundaries of the place and activity locations, and the idea that there is

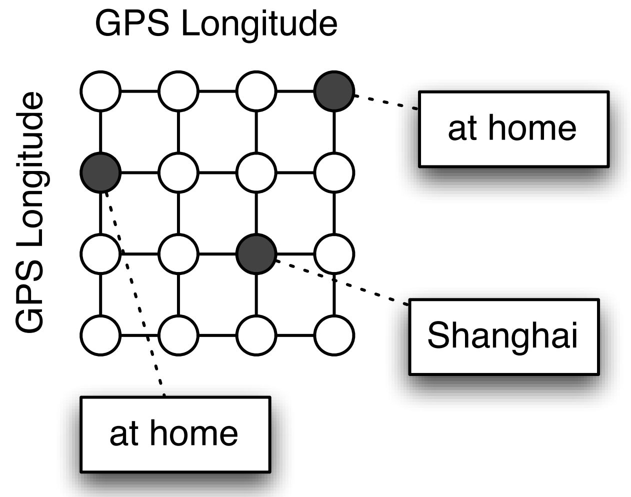

Figure 4: Markov Network based on a GPS model labelled with place observations.

actually just one correct label for a position. But one advantage of this approach is that solving the aposteriori distribution on a given cell should produce a range of reasonable values that, although not modelled as such, can be considered as multiple options for labelling that grid cell. Figure 4 shows a graphical representation of the model instantiated with place labels.

This model has many commonalities with image segmentation in which each cell in the planar belief net is a pixel rather than a physical position. Unlike image segmentation however, there isn't an observation at every GPS position.

To solve this we use can use a standard Markov Network formulation:

$$P(\vec{y}|\vec{O}) = \frac{1}{Z(x)} \prod_{c} \prod_{c\prime} \phi_{c,c\prime}(\vec{y}, \vec{O})$$

$$Z(x) = \sum_{\vec{y}} P(\vec{y}|\vec{O})$$

In this formulation however, the effectiveness is in the definition of the potential function, $\phi_{c,c\prime}$. There are some desiderata that we need to include in our potential function:

- Privacy: It is important that one particular person labelling a location in a sensitive way not cause the solution to reveal that label. For example, labelling a spot, "Bob's house" based on one person's entry would be inappropriate. Modelling this will require the introduction of relational features which aggregate over the number of individuals postulating a given label.
- GPS Noise: GPS has inherent error which needs to be compensated for by allowing labels observed in one location to influence labels in another location. This can be handled by the existing correlated edges.
- Consensus not Frequency: Although one person may label a place 100 times with the same place label, this

Figure 5: Classification of a 15 km by 15 km region of Southern California modelled with the simple GPS model .

should not cause a competing label agreed on by 50 individuals one time each to be ignored. In fact the opposite is true. We wish consensus to dominate frequency. This will also require relational features to be added to the model.

- User Idiosyncrasy: One of the ways in which people use the Nomatic tool is to label the place and activity that they are about to go to and/or do. So there should also be smoothing across individual trajectories. This requires adding links to the network which correspond to a single users trace through the world.

- GPS Lag: In addition to GPS noise, GPS devices are frequently slow to sync with satellites when emerging from indoor locations. Allowing smoothing of labelling across temporal trajectories is also important.

- Curated Entries: Because there are going to be many locations which are not going to be initialized with Nomatic it is worthwhile to include a mechanism for supporting curated labeling of entries. These are like user observations, but are immune form privacy and consensus concerns.

Figure 5 shows the results of modelling the simple GPS model applied to our data. In this 15km square area, most likely place names are each given a unique color. More variation is apparent in regions with denser data collection.

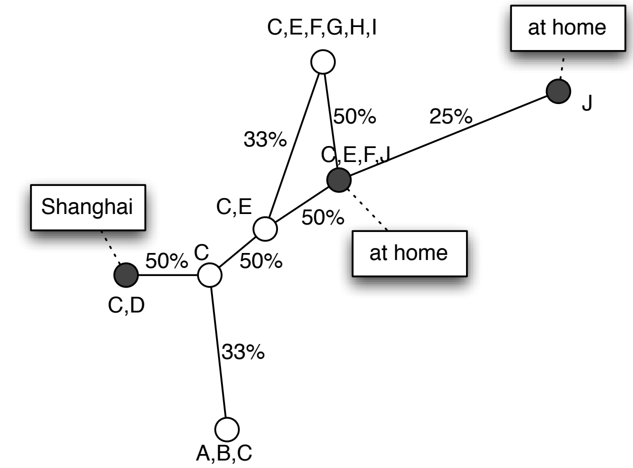

Figure 6: Markov Network based on an AP wifi model high edge weights are characterized by physical proximity.

WiFI Access Point Modelling

Because not all users have GPS devices, an alternative method of modelling the world is by treating combinations of wifi APs as unique locations. Every wifi access point is associated with a unique MAC address (*e.g.*, 00:19:07:d4:14:00), when Nomatic makes an entry in the community history it records all wifi access points that are detected from the platform at the given time. By treating each combination of detected access points as a location, and by connecting two access points when any of the detected access points are in common, and by labelling the edges with weights which are the percentage of overlap it is possible to get a graph which has characteristics of the physical layout of a wifi network. See figure 6 for example. Edge weights are calculated according to this formula:

$$w(\vec{AP}_1, \vec{AP}_2) = min\left(\frac{|\vec{AP}_1 \cup \vec{AP}_2|}{|\vec{AP}_1|}, \frac{|\vec{AP}_1 \cup \vec{AP}_2|}{|\vec{AP}_2|}\right)$$

All of the non-GPS specific desiderata apply to modelling this network as well. The primary difference between these two approaches is the non-regularity of the AP model and the density of observations in the AP model. It is the rare case that a node in the AP model wouldn't have an observation, whereas in the GPS case observations are much sparser.

Hybrid Approaches

Since wifi APs are physically located, it is possible to create a hybrid of these two models in which wifi access points are tied to specific geographic positions when possible. This corresponds to having the wifi model overlay the physical model and having the two models pinned together at places where access points locations are known. The process of pinning wifi APS to a physical location is known as wardriving.

Conclusion

In this paper we have introduced a data set that lends itself to global modelling of place and activity label priors. It is not an abstract set of labels however, it is grounded in a very specific situated use of labelling that benefits the users who participate in the system.

By modelling the physical situation of the user and incorporating features which help to accomodate the unique ways in which physical data is collected in our system we hope to develop global models of human behavior.

Acknowledgments

The authors wish to acknowledge the generous support of the U.S. National Science Foundation through grant number HCC: 0713562.

References

[Boger et al. 2006] Boger, J.; Hoey, J.; Poupart, P.; Boutilier, C.; Fernie, G.; and Mihailidis, A. 2006. A planning system based on markov decision processes to guide people with dementia through activities of daily living. *IEEE Transactions on Information Technology in Biomedicine* 10(2):323–333.

[Choudhury and Pentland 2003] Choudhury, T., and Pentland, A. 2003. Sensing and modeling human networks using the sociometer. In Feiner and Mizell (2003), 216–222.

[2005] Choudhury, T.; Lester, J.; and Borriello, G. 2005. A hybrid discriminative/generative approach for modeling human activities. In Kaelbling, L. P., and Saffiotti, A., eds., *IJCAI*. Edinburgh, UK: Morgan-Kaufmann Publishers.

[2002] Colbry, D.; Peintner, B.; and Pollack, M. 2002. Execution Monitoring with Quantitative Temporal Bayesian Networks. In *6th Intl. Conf. on AI Planning and Scheduling*.

[2003] Feiner, S., and Mizell, D., eds. 2003. *7th International Symposium on Wearable Computers (ISWC 2003), 21-23 October 2003, White Plains, NY, USA*. IEEE Computer Society.

[2003] Fitzpatrick, P., and Kemp, C. 2003. Shoes as a Platform for Vision. In Feiner and Mizell (2003), 231–234.

[2003] Hightower, J. 2003. From position to place. In *Proceedings of The 2003 Workshop on Location-Aware Computing*, 10–12. part of the 2003 Ubiquitous Computing Conference.

[2002] Intille, S. S.; Kukla, C.; and Ma, X. 2002. Eliciting user preferences using image-based experience sampling and reflection. In Terveen, L. G., and Wixon, D. R., eds., *CHI Extended Abstracts*, 738–739. ACM.

[1999] Jebara, T., and Pentland, A. 1999. Action reaction learning: Automatic visual analysis and synthesis of interactive behaviour. In *ICVS '99: Proc. of the First Intl. Conf. on Computer Vision Systems*, 273–292. London, UK: Springer-Verlag.

[2004] K. Van Laerhoven, K., and Gellersen, H. 2004. Spine Versus Porcupine: A Study in Distributed Wearable Activity Recognition. volume 1, 142–149. IEEE.

[1986] Kautz, H. A., and Allen, J. F. 1986. Generalized Plan Recognition. In *AAAI*, 32–37.

[2004] Kern, N.; Antifakos, S.; Schiele, B.; and Schwaninger, A. 2004. A model for human interruptability: Experimental evaluation and automatic estimation from wearable sensors. In Smith and Thomas (2004), 158–165.

[2005] Patterson, D. J.; Fox, D.; Kautz, H. A.; and Philipose, M. 2005. Fine-grained activity recognition by aggregating abstract object usage. In Mase, K., and Rhodes, B., eds., *ISWC*, 44–51. Osaka, Japan: IEEE Computer Society.

[2006] Patterson, D. J.; Ding, X.; and Noack, N. 2006. Nomatic: Location by, for, and of crowds. In Hazas, M.; Krumm, J.; and Strang, T., eds., *Location- and Context-Awareness, Second International Workshop, LoCA 2006, Dublin, Ireland, May 10-11, 2006, Proceedings*, volume 3987 of *Lecture Notes in Computer Science*, 186–203. Springer.

[2004] Perkowitz, M.; Philipose, M.; Fishkin, K. P.; and Patterson, D. J. 2004. Mining models of human activities from the web. In Feldman, S. I.; Uretsky, M.; Najork, M.; and Wills, C. E., eds., *WWW*, 573–582. ACM.

[2004] Philipose, M.; Fishkin, K. P.; Perkowitz, M.; Patterson, D. J.; Hahnel, D.; Fox, D.; and Kautz, H. 2004. Inferring activities from interactions with objects. *IEEE Pervasive Computing: Mobile and Ubiquitous Systems* 3(4):50–57.

[2004] Shi, Y.; Huang, Y.; Minnen, D.; Bobick, A.; and Essa, I. 2004. Propagation networks for recognition of partially ordered sequential action. In *Proceedings of IEEE CVPR04*.

[2004] Smith, M., and Thomas, B. H., eds. 2004. *8th International Symposium on Wearable Computers (ISWC 2004), 31 October - 3 November 2004, Arlington, VA, USA*. IEEE Computer Society.

[2003] Stäger, M.; Lukowicz, P.; Perera, N.; von Büren, T.; Tröster, G.; and Starner, T. 2003. Soundbutton: Design of a low power wearable audio classification system. In Feiner and Mizell (2003), 12–17.

[2004] Stäger, M.; Lukowicz, P.; and Tröster, G. 2004. Implementation and evaluation of a low-power sound-based user activity recognition system. In Smith and Thomas (2004), 138–141.

[1983] Wilensky, R. 1983. *Planning and Understanding: A Computational Approach to Human Reasoning*. Reading, MA: Addison-Wesley.