Learning and Predicting Multimodal Daily Life Patterns from Cell Phones

Katayoun Farrahi and Daniel Gatica-Perez
Idiap Research Institute, Martigny, Switzerland
Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
kfarrahi@idiap.ch, gatica@idiap.ch

ABSTRACT

In this paper, we investigate the multimodal nature of cell phone data in terms of discovering recurrent and rich patterns in people's lives. We present a method that can discover routines from multiple modalities (location and proximity) jointly modeled, and that uses these informative routines to predict unlabeled or missing data. Using a joint representation of location and proximity data over approximately 10 months of 97 individuals' lives, Latent Dirichlet Allocation is applied for the unsupervised learning of topics describing people's most common locations jointly with the most common types of interactions at these locations. We further successfully predict where and with how many other individuals users will be, for people with both highly and lowly varying lifestyles.

Categories and Subject Descriptors: I.5.2 [Design Methodology]: Pattern analysis

General Terms: Human Factors.

1. INTRODUCTION

Cell phones are rapidly emerging as the ultimate multimodal sensor of human dynamics [5]. Equipped with GPS, Bluetooth, accelerometers, cameras, and microphones, current phones have the potential of tracing human activities at scales previously unattainable and of enabling the design of new human-centered applications related to people's daily life, thus opening a whole scope of problems in multimodal integration and ubiquitous computing [2, 8, 9].

Two fundamental problems in this domain relate to routine modeling: how to *discover* recurrent patterns in a person's life from multimodal data like proximity, location, and motion, and how to *predict*, based on the knowledge of a person's routines, her most likely routines at any given time. On one hand, pattern discovery via unsupervised learning is often a necessity, given the potentially large number of relevant routine patterns of an entire population and the huge amount of unlabeled data that can be recorded with a phone over time [3, 4]. On the other hand, predictions from aggregated user observations are, arguably, some of the most useful outcomes of routine modeling, by inferring both where and with whom a user would most likely be in the future (for anticipation) or would most likely have been in the past (for cases of missing data).

While recent works have started to analyze both problems from location or proximity data - discovery and prediction in [3], discovery in [4] - one aspect that has not been investigated in depth is the role of multimodal integration in large-scale routine analysis. More specifically, how does the joint use of multiple modalities (e.g. location and proximity to others) enhance the understanding of a person's routines, and how can this be efficiently represented and automatically inferred? Proximity to known people (as a coarse approximation of face-to-face interaction) adds a rich element of social context that is very useful to complement or disambiguate many situations in daily life. For instance, being at home alone and with a large group having a party represent entirely different social situations, that would be nevertheless identical from the sole perspective of location. Such finer descriptions of routines based on multiple cues are clearly important to characterize users and their habits.

This paper presents an approach for large-scale unsupervised learning and prediction of people routines through the joint modeling of human location and proximity interactions. Our work has three contributions. First, extending our previous work [4], we propose a multimodal representation that integrates location and proximity to characterize a person's daily life in a simple yet robust way. Second, we show that a probabilistic topic model approach for routine mining results in the discovery of patterns that are not only meaningful but also complementary, informing about a person's life better than location alone or proximity alone. Finally, we show that the discovered topics can be further used for prediction purposes (i.e. inferring missing or future bits of a person's life), present a method to do so, and show promising performance on a massive and challenging dataset involving 97 people and thousands of days of data.

2. MULTIMODAL FRAMEWORK

We use the Reality Mining (RM) dataset [3] for which the activities of 97 students and staff at MIT were recorded by Nokia 6600 smart phones over the 2004-2005 academic year. Given a day in the life of a person in terms of where they go and the number of people within the group they are in proximity with, our goal is to discover routines from large-scale multimodal phone data. Further, we use the combined lo-

ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.
Copyright 2009 ACM 978-1-60558-772-1/09/11 ...$10.00.

cation and proximity routines discovered to predict missing location and proximity data. Following [4], we represent a day in the life of a user in terms of where they are over a 90-minute time interval as well as the number of people they are with during this time interval within the RM population, forming a joint location-proximity data representation described next.

2.1 Joint Location-Proximity Representation

The joint location-proximity data representation is based on the concatenation of data corresponding to users' location, proximity, and a timeslot indicating a coarse measure of the time of day for which this data is measured. The details follow.

Location Representation: Following Eagle et al [3], a given individual's locations (given by cell towers) is represented over the course of a day by first simplifying all possible locations into 4 categories, work (W), home (H), out (O), and no reception (N). W are the MIT work premises, H are the homes of individuals, and O are towers that are not H or W. N is a label used if there is missing data for a person for a given time. The basic idea for the location word representation, which is taken from our previous work [4], is to assign a single location label (H,W,O,N) for each 30−minute time interval of a user's day, resulting in 48 location labels for each user and each day. Then, 3 consecutive 30-minute labels are taken to obtain location transition information over a 1.5 hour period in a day. These 1.5-hour intervals are overlapping, resulting in 48 x 1.5-hour 3-label location sequences in a day.

Timeslot Division: Each day is divided into 8 timeslots as follows: 0-7am (1), 7-9am (2), 9-11am (3), 11am-2pm (4), 2-5pm (5), 5-7pm (6), 7-9pm (7), 9-12pm (8). This timeslot is concatenated to the 1.5 hour location label sequences to form the location words used in [4].

Proximity Representation: For proximity data, we only consider proximity with people in the RM group. Proximity in general could be considered, though proximity with laptops and computers is also recorded in the data and is difficult to distinguish from mobile phones. We quantize the number of proximate people into 4 prototypical groups: user is alone, dyad (1 person in proximity), small group (2-4 people in proximity), large group (5 or more people in proximity). The group sizes are motivated by research in social science that has traditionally analyzed dyads, small groups, and large groups as separate categories, as they present distinct dynamics.

A day in a user's life is finally represented as a bag of words, where a word is a location word, concatenated with a proximity group and a timeslot.

2.2 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is an unsupervised probabilistic generative model that was initially developed to characterize text documents, but can be extended to other collections of discrete data [1]. A *word* is a basic unit of discrete data defined by an item for a vocabulary of size V. A *document* is a bag of N words, and a corpus is a collection of M documents. Each document is viewed as a mixture of topics, where topics are distributions over words. The probability of a given word w_t assuming T topics is given by: $p(w_t) = \sum_{t=1}^{T} p(w_t|z_t)p(z_t)$, where z_t is a latent variable indicating the topics from which the t^{th} word was drawn.

The objective of LDA inference is to determine the word distribution $p(w|z = t) = \phi_w^{(t)}$ for each topic t and the topic distribution $p(z = t) = \theta_t^{(d)}$ for each document d. We use the approximation derived in [6] based on Gibbs sampling. In LDA, $p(\theta)$ and $p(\phi)$ are assumed to have Dirichlet distributions with hyperparameters α and β, respectively. The Gibbs sampler is used since the estimation problem of maximizing $p(w|\phi, \alpha) = \int p(w|\phi, \theta)p(\theta|\alpha)d\theta$, is intractable. The Gibbs sampler results in

$$\phi_t^{(w)} = \frac{n_t^{(w)} + \beta}{n_t + V\beta}, \; \theta_d^{(t)} = \frac{n_d^{(t)} + \alpha}{n_d + T\alpha}, \tag{1}$$

where $n_t^{(w)}$ and $n_d^{(t)}$ are the number of times word w and document d have been assigned to topic t, respectively, and $n_t = \sum_{w=1}^{V} n_t^{(w)}$ and $n_d = \sum_{t=1}^{V} n_d^{(t)}$. In our work, documents are days and words are defined in Section 2.1. We use LDA for two tasks:

Routine Discovery: We recently used LDA discover location routines [4]. Here, we propose to extend this use to handle multimodal data, expecting that topics will capture joint patterns of location and proximity that help disambiguate relevant cases (e.g. discriminating between a person at work alone and in a group).

Predicting Behavior: LDA is also used for the prediction of missing words in a document (i.e. the prediction of users' joint patterns of location and proximity for certain timeslots). To achieve prediction, LDA inference is run on the unseen documents with missing bits. This results in matches between an unseen document and known documents using the estimation of the posterior distribution of topics given the unseen documents' words $p(z|w)$. The resulting topics are ranked according to $p(d|z)$ and the top 3 topics z_{top} are selected for potential use in prediction. The most probable word's timeslot for topics z_{top} is compared to the missing sequence's timeslot. We pick the topic z_{top} whose top word's timeslot is the closest to the timeslot in the document whose missing words will be predicted. We also ensure that $p(d|z_{top}) > Th$ where Th is chosen to ensure the document is characterized by the topic with high enough probability. The result is a single topic which best describes the missing data over the timeslot. To fill in the missing location and proximity words, we replace the missing labels with those of the top document for the mostly likely topic selected.

3. EXPERIMENTS AND RESULTS

We experimented with all of the 97 individuals in the dataset and days ranging from 18.07.2004 to 05.05.2005, encompassing 291 consecutive days thus extending our previous work [4] who only considered 30 users. This subset of days was chosen since these are the days for which proximity data is mostly available. Days with entirely no reception for location were not considered since they contain no useful information for proximity either. The LDA model used for joint location-proximity routine discovery had $T = 100$ topics. Heuristic methods were used to obtain T, but roughly speaking, a small value of T will produce coarse routines, whereas a large T will be much more specialized.

3.1 Joint Location-Proximity Routines

The fusion of proximity and location data enables the discovery of more detailed information regarding this group of

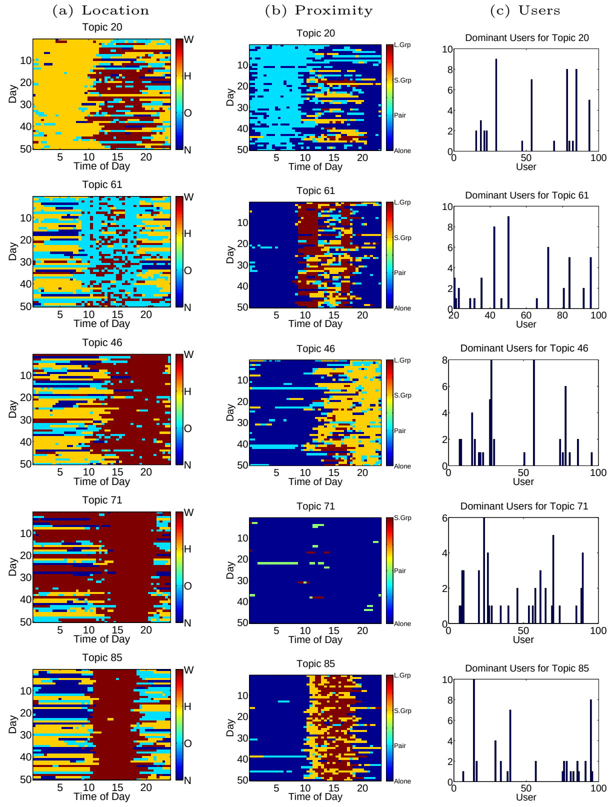

Figure 1: Ranked days (i.e. documents) for all topics by $p(d|z_i)$, showing the top 50 days' location data (a) and proximity data (b) for a given topic, z_i. (c) A histogram of the users whose days ranked in the top 50 for topic z_i. Note the colorbars for the location figures (a) indicating the W, H, O, and N locations and for the proximity figures (b) indicating a large group, small group, pair, or alone.

MIT users' daily lives compared to single modalities. A short summary of the entire corpus is presented below, but a more in depth analysis of the results in the form of a video demo can be found at www.idiap.ch/ kfarrahi/MMDemo/results.wmv.

-*Home routines and proximity:* Most of the home routines discovered occurred for users alone (not in proximity with anyone from the group). Only 2 out of the 20 home routines discovered dominated for a pair of users in proximity. No home routines occurred for small or large groups which suggests that people did not socialize within the population at home.

-*Work routines and proximity:* Most of the routines discovered with proximity interactions occurred at work locations. There are 17 work routines, and 13 of them occur with proximity patterns. Routines at work were discovered for all four proximity groups (users alone, in dyads, small and large groups), which indicates that all these types of interactions occur frequently.

-*Morning routines and proximity:* Only 3 out of 100 topics had a proximity interaction in the morning (before 10am), and all 3 of these routines occur for pairs of users and never for groups. People interacting in the morning seems to be relatively sparse.

-*Day-time routines and proximity:* Approximately 20 topics characterize user interactions throughout the day (10am-7pm). The interactions include pairs of users, as well as small and large groups.

-*Evening routines and proximity:* 7 topics characterize group interactions in the evenings (7pm-midnight). These occur for pairs of users, and small as well as large groups.

Specific topics illustrating the types of joint routines discovered are visualized in Figure 1. We have illustrated results for selected topics $z_i = 20, 61, 46, 71, 85$, where column (a) visualizes the location data, column (b) the proximity data, and column (c) the user statistics. For a selected topic z_i, the top 50 documents are ranked according to $p(d|z_i)$ and their location data (a) and proximity data (b) are visualized. Further, a histogram for the users whose days ranked in the top 50 documents is shown in (c). A summary of the routines discovered plotted in Figure 1 is:

-**Topic 20**: At home in the morning in proximity to someone in the group. This routine occurs for 5 specific individuals mostly (as seen in the plot "Dominant Users for Topic 20").

-**Topic 61**: Out roughly from 10am-7pm as a large group with a "break" occurring in between for several hours in work locations. This might correspond to days with courses in the morning and afternoon for several students which are held off their main work environment.

-**Topic 46**: At work in the afternoon until late in the evening with a small group of people. This corresponds to typical graduate student behaviour.

-**Topic 71**: At work non-stop from roughly 1pm-8pm alone. This routine dominates for many individuals.

-**Topic 85**: At work from 10am-7pm in small to large groups. This occurs most frequently for 3 specific individuals, who seem to have structured working hours.

3.2 Behavior Prediction

The purpose of behavior prediction is to apply LDA in order to predict unobserved location and proximity data for a timeslot of a user's day. For experiments, we decided to separate users based on the entropy of users' lifestyles [3]. In our work, we use the Author Topic Model (ATM) to separate users based on the entropy of their location routines. In [4] we used the ATM for routine discovery but not for entropy characterization. Entropy is computed on the probability of topics given authors $p(z|a)$, where an author a is a user, location words are the same, and a document is a day. All of the users in the dataset are ranked according to their entropies. We set two thresholds for high and low entropy which gave 10 users in each case. We randomly picked 5 for each class (high and low).

For each of the 10 users picked, 20 days were randomly selected, from days with at least one proximity interaction (i.e. at least one word over the entire day contains an interaction). This set of days was used to form the test set from which we remove words to generate data with missing sequences to predict. For each day, the words of a given timeslot were removed to form a day from which the method predicts the missing sequence, thus generating 8 days, each with one timeslot's words missing. The resulting dataset from which we predict missing sequences contains 10 users, each with 160 days = 1600 documents for testing. Thus, for each user there are 160 documents for testing, and each timeslot contains 200 documents for testing.

For each document, there is one timeslot with missing location and proximity labels. The *location error* is the number of incorrect labels divided by the total number of labels to be predicted in the given timeslot. For instance, documents with timeslot 1 missing have 14 location labels to be predicted since it occurs from 0-7am. The *proximity error* is the average number of people wrongly predicted for each word in the timeslot. More specifically, if the predicted group (alone, dyad, small group, large group) is correct then there is no error. If the predicted group is incorrect, then we predict the minimum number of possible people in the group (alone=1, dyad=2, small group=3, large group=5) and compute the difference with the actual number of people in proximity. For example, if there are 10 people in proximity and we predict a small group, then we assume 3 people are in proximity. If this incorrect prediction occurs over the 14 words in timeslot 1, then the average proximity error is 7.

The location and proximity errors are computed over users and timeslots in Figure 2. We can see the average errors as a function of the user for location in Fig 2(a) and proximity in Fig 2(b). Users 1-5 (in blue) have low entropy and 6-10 (in red) have high entropy. In Fig 2(a), the bar shows the error which would be obtained if all labels would be labeled as 'W', which is the most frequently occurring location label in the test set. Interestingly, low-entropy users have lower error in the prediction of location labels than high-entropy users. In Fig 2(b) we plot the proximity error. In the best (resp. worst) case, the predicted number of people in proximity is incorrect by, on average, 0.5 (resp. 1.5) people. In this case, low entropy users do not necessarily have lower prediction errors in proximity than high entropy users. At the same time, for these results, entropy was computed on users' location, and these results show that users with predictable location patterns do not necessarily have predictable proximity patterns.

In Figures 2(c) and (d), we plot the average errors as a function of timeslot for both high and low entropy users for location (Fig 2(c)) and proximity (Fig 2(d)). We can see in Fig 2(c) that for every timeslot, high entropy users are harder to predict (have higher errors) than low entropy users. Also, for timeslots 1 and 2, low entropy users correspond to much better performance than high entropy users. The worst performance occurs for timeslots 6, 7, and 8, especially for high entropy users, thus we can conclude the prediction of location is most difficult in the evenings. The error in proximity prediction as a function of timeslot, in Fig 2(d), is again not highly correlated with the location entropy of a user. The prediction in proximity has the highest error in timeslot 5, corresponding to 2-5pm and the lowest error in the mornings. In the worst case, the proximity error is less than 2 people on average.

4. CONCLUSIONS

Our method successfully discovers recurrent patterns in people's lives from multimodal data and can use the discovered routines for data prediction, estimating location and proximity data of users with varying location entropy. In future work, the methodology for data prediction will be further optimized to include prediction on varying timescales, to predict one data modality source given another (for example, to predict a user's location given the time of day and their interactions), and to consider proximity with all bluetooth devices including laptops, computers, and anonymous cell phones.

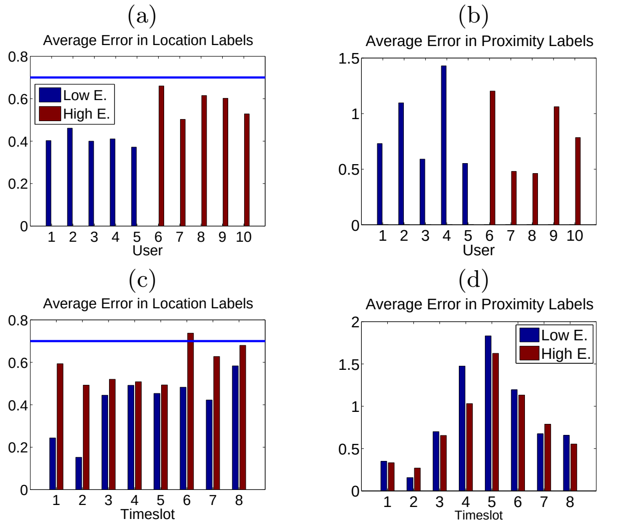

Figure 2: (a) Average location error for prediction as a function of users, where low entropy users are labeled 'Low E' and high entropy users 'High E'. (b) Average proximity error for prediction as a function of users. The average error in (c) location prediction, and (d) proximity prediction, as a function of timeslot for low and high entropy users.

Acknowledgments

This research has been supported by the Swiss National Science Foundation through MULTI. We thank Nathan Eagle (MIT) for sharing the data.

5. REFERENCES

[1] D. Blei, A. Ng and M. Jordan. "Latent Dirichlet Allocation," *Journal of Machine Learning Research 3*, 2003.

[2] T. Choudhury, M. Philipose, D. Wyatt and J. Lester. "Towards activity databases: Using sensors and statistical models to summarize people's lives," *IEEE Data Eng. Bull:49-58*, 2006.

[3] N. Eagle and A. Pentland. "Eigenbehaviors: Identifying Structure in Routine," *Behavioral Ecology and Sociobiology (in submission)*, 2007.

[4] K. Farrahi and D. Gatica-Perez. "What Did You Do Today? Discovering Daily Routines from Large-Scale Mobile Data," *Proc. ACM Int. Conf. on Multimedia (MM)*, Vancouver, 2008.

[5] M.C. Gonzalez, A. Cesar and A.L. Barabasi. "Understanding Individual Human Mobility Patterns" *Nature 453(7196):779-782*, 2008.

[6] T.L. Griffiths and M. Steyvers. "Finding Scientific Topics," *PNAS 101:5228-5235*, 2004.

[7] Sense Networks, "http://www.sensenetworks.com".

[8] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. "Computational Social Science," *Science*, Feb. 2009.

[9] H. Lu, W. Pan, N. Lane, T. Choudhury, and A. Campbell. "SoundSense: scalable sound sensing for people-centric applications on mobile phones," *Mobisys '09: Proceedings of the 7th international conference on Mobile systems, applications, and services:165-178*, 2009.