Bayesian Nonparametric Modeling of User Activities

Yin Zhu[1,2], Yuki Arase[2], Xing Xie[2] and Qiang Yang[1]
Hong Kong University of Science and Technology
Microsoft Research Asia
{yinz,qyang}@cse.ust.hk, {yukiar,xingx}@microsoft.com

ABSTRACT
Human activity modeling is becoming more and more important in ubiquitous computing as it builds a foundation for higher-level applications in areas such as e-health and activity recommendation systems. Many existing works in this area focus on recognizing a pre-defined set of activities using some devices in the supervised learning setting, however, it is hard to define activities and label sensor data, especially for a new environment. In this note we aim to recognize activities in an unsupervised way – segment activity sensor reading sequence and group the segments into meaningful categories by leveraging Sticky HDP-HMM. We have conducted experiments on a sensor dataset collected in an office area using a smartphone and the result shows that our method frees annotation process and renders good activity recognition result.

Author Keywords
Activity Recognition, Unsupervised Learning, Sticky HDP-HMM.

ACM Classification Keywords
H.1.1. Systems and Information Theory: Information theory.

General Terms
Algorithms, Experimentation.

INTRODUCTION

Activity recognition using sensors [7] is an important building block in ubiquitous computing systems. Many existing activity recognition works focus on recognizing activities in a supervised setting – labeled data are collected from some sensor devices and a pre-defined set of activities could be recognized by the model learned from the labeled data. Although supervised method is more accurate for prediction accuracy, it needs labeling effort and does not scale. Thus unsupervised modeling for activity recognition becomes a demanding research topic in ubiquitous computing area.

TDMA'11, September 18, 2011, Beijing, China.

In this note, we study a new problem in this direction: Segment the sensor reading into meaningful segments and count the number of different types of them at the same time. This work can greatly free the data labeling process for activity recognition. Figure 1(a) illustrates this problem and the motivation behind. The signal strength [1] of an accelerometer in 46 seconds with 5Hz sampling frequency is recorded when the smartphone user carries the phone to the water machine, get some hot water, go back with the water and put the phone on the desk then sit down. Figure 1(b) and 1(c) are segmentation results generated by our approach with different parameters. Figure 1(b) shows a segmentation of the strength stream with different colors representing different categories of activities. We can see the clear correspondence between the true activity labels and the segmentation colors. A bonus from this setting is that the number of different activities is automatically obtained during the segmentation process by grouping similar segments into clusters. Figure 1(c) shows another feature of our work, i.e. segmentation at different granularity levels. Because moving with an empty cup and moving with a cup of hot water share similar patterns, they are merged into the same activity cluster.

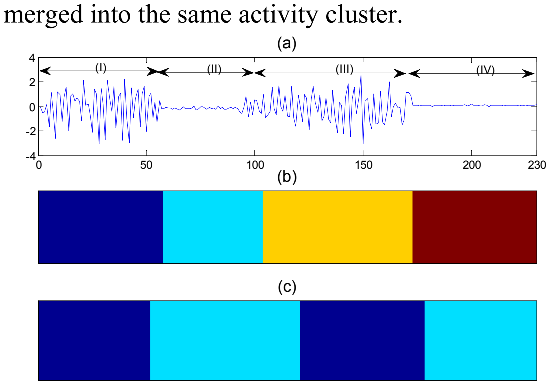

Figure 1. (a) Accelerometer strength in 46 (=230/5Hz) seconds: (I) go to the water place; (II) get water; (III) go back with hot water; (IV) put the phone on desk. (b) Segmentation at detailed granularity level (κ=0.89). (c) Segmentation at rough granularity level (κ =4.00).

Aside from good metrics of unsupervised learning such as reduced labeling effort, our work in particular can find some novel activities that are not noticed by people before.

[1] Defined as $\text{strength} = \sqrt{x^2 + y^2 + z^2} - 9.8\text{m/s}^2$, where x, y and z are velocities along the three axes of the accelerometer.

Sometimes the aim of the application is to find novel or abnormal activities [5]. In this case, pre-defining activities in the supervised setting is less interesting. If we limit the number of the activities we can detect, how can we find new ones? In the experiments part, we will show that sub-activities of "moving around" activity can be differentiated by the types of the segments.

There are previous works focusing on automatic segmentation for sensor data, e.g. [13]; there are also works on unsupervised modeling of the activities, e.g. activity topic modeling in [6]. Our work can be viewed as the combination of these two tasks together – segmentation and assigning topic/cluster to segments at the same time. With a method of obtaining the number of activity categories automatically, our work is advantageous over that of Huynh *el. al* [6] which assumes a known number of topics beforehand.

To solve this challenging problem, we leverage the recent advances in nonparametric Bayesian models and apply Sticky HDP-HMM [3] as our pattern finding algorithm. Sticky HDP-HMM extends the original HDP-HMM by encouraging self-transitions in the hidden states so that it avoids frequent transitions between states in the original HDP-HMM. A review of this model and the application to activity recognition is described in the next section.

Our contributions of this note are three-folds:

- By leveraging the nonparametric Bayesian model, Sticky HDP-HMM, we can automatically divide raw sensor values into meaningful segments.
- The segmentation can be done at different granularity levels and the number of activity categories can also be obtained from the data.
- Experiments in an office area validate our methodology.

STICKY HDP-HMM BASED ACTIVITY MODELING

Sticky HDP-HMM Background

Hierarchical Dirchlet Process [12] can be viewed as a kind of generative (infinite) mixture models where each hidden topic generates observations using some probability distribution. As a nonparametric Bayesian model, HDP can learn the number of topics/clusters in the data automatically while parametric Bayesian methods, e.g. Latent Dirichlet Allocation (LDA) [1] and Probabilistic Latent Semantic Analysis (PLSA) [4], need this number to be provided beforehand. One important application of HDP is the extension on Hidden Markov Models (HMMs). An HDP formulation can decide the right number of states for the Hidden Markov Model (HMM) from its posterior density function on the appropriate number of mixture components, to some extent, the number of states in HMM can go to infinite if necessary. However, the states that have actually emitted observations are finite and the number is automatically learned from the data while in HMM the number of states should be provided. Besides, HDP-HMM learns the appropriate degree of sharing of data across data sets through the sharing of mixture components.

The core mathematical component in DHP is Dirichlet process, denoted by $\mathrm{DP}(\gamma, H)$, which is a distribution over countably infinite random measures:

$$G_0|\gamma, H \sim DP(\gamma, H)$$

$$\mathrm{G}_0(\theta) = \sum_{k=1}^{\infty} \beta_k \delta\left(\theta - \theta_k\right), \qquad \theta_k \sim H, \beta \sim GEM(\gamma)$$

where $GEM(\cdot)$ stands for stick-breaking process as follows:

$$\beta_k' \sim Beta(1, \gamma)$$

$$\beta_k = \beta_k' \prod_{l=1}^{k-1} (1 - \beta_l'), \quad k = 1,2, \dots$$

HMM [7] can be viewed as a doubly stochastic Markov chain and is essentially a dynamic variant of a finite mixture model. Therefore, by replacing the finite mixtures with a Dirichlet process, we can complete the design of HDP-HMM (as shown in Figure 2, but without the parameter κ):

$$\gamma \sim GEM(\eta)$$

$$\pi_k \sim DP(\alpha, \gamma), k = 1, \dots, \infty$$

$$s_t \sim Multinomial\left(\pi_{s_{t-1}}\right), t = 1, \dots, T, s_0 = 1$$

$$\phi_k \sim H, \qquad k = 1, \dots, \infty$$

$$y_t \sim F\left(\phi_{s_t}\right), \qquad t = 1, \dots, T$$

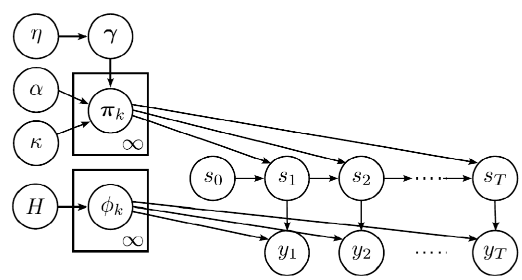

Figure 2. The graphical model for Sticky HDP-HMM. The parameter κ controls the self-transition rate between states. When $\kappa = 0$, the model degenerates into the original HDP-HMM [3].

While HDP-HMM is theoretically solid for model averaging; however, one major defect of HDP-HMM in practice is that it inadequately models the temporal persistence of states – the states transitions could be too frequent and the number of hidden states goes big. In activity recognition, an activity of some semantic meaning usually lasts for some time. If we use one hidden state to represent one activity, there may occur situations where an activity is split into several sub-activities which do not have concrete meanings. Sticky HDP-HMM shown in Figure 2 augments HDP-HMM with a parameter κ , which effectively controls the switching rate among the hidden states in HDP-HMM. By introducing the self-transition parameter κ, the new distribution for π_k is:

$$\pi_k |\ \alpha, \kappa, \beta \sim DP\left(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha + \kappa}\right)$$

where when $\kappa = 0$, the Sticky HDP-HMM degenerates to the original HDP-HMM. Chinese restaurant process (CRP) [4] metaphor can explain κ intuitively: Introduce the concept of "loyal customers" into the CRF, each restaurant now has a specialty dish that has the same index as the restaurant. This dish is served everywhere but is more popular in its namesake restaurant. In other words, each restaurant now has a specific rating of the buffet line that puts more weight on the specialty dish.

Due to the space limit, we omit the MCMC inference method and other details of Sticky HDP-HMM; interested readers are referred to [3] for details.

Activity Recognition Modeling

One previous successful application of Sticky HDP-HMM is speaker diarization [3], the problem of counting the number of speakers in a meeting where the original HDP-HMM tends to over segment the recorded audio data. The activity recognition problem shares the same property with speaker diarization – we automatically segment the sensor data while counting how many different kinds of segments therein.

EXPERIMENT

Data Collection

We have collected the sensor data from three subjects (A, B and C) who are students in our research lab in an office area. Each subject carries a smartphone running Windows Mobile 6.5 and the phone is installed with a small application collecting accelerometer and WiFi access points' information. The sampling frequency for accelerometer is 5Hz and the scanning frequency for nearby WiFi access points is ½ Hz (i.e. scanning every 2 seconds). We divide the data into small windows of 2 seconds so that every window has one record for nearby WiFi access points and about 10 accelerometer readings. Each subject carries the smartphone for one working day. The working environment is within our institution building which has WiFi signals everywhere and also has several restaurants in it. To provide the ground truth to evaluate the segmentation quality, we have manually labeled the segments into 5 categories: *work at seat*, *meeting*, *seminar*, *eating* and *moving around*.

We extract 6 features from the accelerometer strength in a window: mean, average, absolute mean, energy, sum of FFT coefficients, weighted sum of FFT coefficients. These features are commonly used for accelerometer based application, e.g. transportation mode detection [9].

We use the raw signal strength from WiFi access points as other features. WiFi signals are used to help indicate the location of the subject [11].

Because WiFi signal strength usually varies between -40db to -100db and to make the feature vector balanced between WiFi signals and accelerometers, we further normalize the values of these 6 accelerometer features to zero mean and 100 variance to ensure the feature values are comparable to that of WiFi's.

Evaluation Method

Evaluating the performance of segmentation/clustering has many metrics. We follow the common evaluation metrics for clustering [10] and also use an entropy-based error defined as follows:

$$E(S) = \frac{1}{M}\sum_{i=1}^{M} En\big(L(s_i)\big),$$

where the true segmentation $S = \{s_i\}_{i=1}^{M}$, s_i is an interval set on the sensor stream and M is the total number of segments. Probability function $L(s_i)$ measures the generated segments distribution over the segment s_i. Finally, the error value of segmented result S is the average entropy of the discrete distributions $L(s_i), i = 1,2 \ldots M$.

An intuitive understanding of this entropy-based error is mapping the segment results returned by Sticky HDP-HMM into every true activity segment and check the purity of this mapping.

Experimental Result

Our first experiment is to show the ability of Sticky HDP-HMM in segmenting accelerometer data at different granularity levels. This example is already discussed in the Introduction section as shown in Figure 1. The self-transition rate parameter κ in Sticky HDP-HMM can effectively guide the algorithm to segment sensor data with different granularity levels.

The main experiment is to compare Sticky HDP-HMM with the LDA method used in Huynh *et. al* [6] as a baseline method. For each data sample, LDA outputs a distribution over the topics for it. Each topic in LDA could be viewed as a cluster, and a data sample has a distribution (thus soft-cluster) over topics. We choose the topic with maximum probability to represent the sample. Also because LDA needs a specific parameter K to indicate the number of topics, we try three different K values: 5, 10 and 15 respectively.

Methods	*E (S)*
Sticky HDP-HMM	0.0131
LDA (K=5)	0.0514
LDA (K=10)	0.0767
LDA (K=15)	0.0828

Table 1. Segmentation error *E(s)* for Sticky HDP-HMM and LDA.

Table 1 shows the comparing result for the entropy-based errors of Sticky HDP-HMM and LDA models. Sticky HDP-HMM performs better than various LDAs. To illustrate the performance of the two algorithms, we draw comparing plots for the segmentation results on subject A's data in Figure 3. We can see clearly that the segmentation from Sticky HDP-HMM is more continuous while LDA (K=10) based segmentation has more outliers. This is mainly due to the fact that LDA does not consider any sequential information.

We have also compared the ground truth labeling of subject A's sensor data and the segmentation returned by Sticky HDP-HMM, we find that Sticky HDP-HMM inferred 12 hidden states from the data ($\kappa = 0.5$), which is bigger than the number of manual labels (5). The more number of segment types are mainly on the *moving* activity part, i.e. the Sticky HDP-HMM outputs more detailed segmentation result. For instance, based on the segmentation result and knowledge of the working pattern, going to the printer (between 12:00 – 13:50) and going to the kitchen for fruits (around 15:00) are clearly distinguished; while in subject A's labeling diary, these sub patterns are not differentiated out.

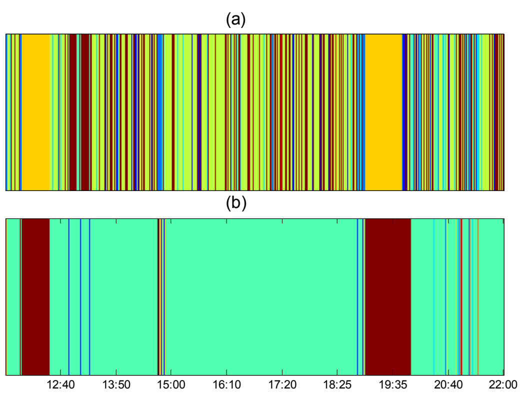

Figure 3. Comparing the segmentation results of LDA (a) and Sticky HDP-HMM (b) for Subject A's one day sensor data. Figure (b) is much smoother than (a).

CONCLUSION AND DISCUSSIONS

We have applied the Sticky HDP-HMM for unsupervised activity modeling from sensor data. The nonparametric nature of this model allows the number of activities to be learned from the data. More importantly, the unsupervised segmentation of the sensor data is close to the true labeling as shown in our experiments. We have also shown that the segmentation could be done at different granularity levels.

In the future, we will try to encode domain knowledge into this kind of unsupervised modeling and generate the semantic label for a sensor segment rather than a category ID in this work. A though empirical study on how κ in Sticky HDP-HMM affects the granularity should also be considered. Another interesting application of work is to use our method as a visualization tool of raw sensor data that assists the user to annotate the sensor data more efficiently.

REFERENCES

1. D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
2. D. Blei, T.L. Griffiths, M.I. Jordan and J.B. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. In *Advances in neural information processing systems*, pages 106-114, 2004.
3. E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky. An hdp-hmm for systems with state persistence. In *ICML*, pages 312–319, 2008.
4. T. Hofmann. Probabilistic latent semantic analysis. In *UAI*, pages 289–296, 1999.
5. D. H. Hu, X.-X. Zhang, J. Yin, V. W. Zheng, and Q. Yang. Abnormal activity recognition based on hdp-hmm models. In *IJCAI*, pages 1715–1720, 2009.
6. T. Huynh, M. Fritz, and B. Schiele. Discovery of activity patterns using topic models. In *UbiComp*, pages 10–19, 2008.
7. Donald J. Patterson, Lin Liao, Dieter Fox, Henry A. Kautz: Inferring High-Level Behavior from Low-Level Sensors. In *UbiComp*, pages 73-89, 2003.
8. L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. In *Proceedings of the IEEE* 77, 257-286, 1989.
9. S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using mobile phones to determine transportation modes. *ACM Trans. Sen. Netw.* 6, 2, Article 13, 2010.
10. J. Wu, H. Xiong, J. Chen. Adapting the Right Measures for K-means Clustering. In *KDD*, pp. 877 - 886, 2009.
11. Q. Yang, S. J. Pan, V. W. Zheng: Estimating Location Using Wi-Fi. *IEEE Intelligent Systems* 23(1): 8-13 (2008).
12. Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei Hierarchical dirichlet processes. *Journal of the American Statistical Association*. 101(476),pages 1566--1581, 2006.
13. J. Yin, D. Shen, Q. Yang, and Z.-N. Li. Activity recognition through goal-based segmentation. In *AAAI*, pages 28–34, 2005.