Constructing an Ecologically-Valid Formal Markovian Model of Human Activity Patterns

Beiyu Lin, Washington State University, beiyu.lin@wsu.edu

***Abstract*—While pervasive computing technologies are becoming mainstream for observing human behavior, few formal models have been theorized of behavior based on collected sensor data. We construct mathematical models that express the nature of human behavior captured by smart home sensor data. We explore the possibility of describing behavior for a sample population as well as population subgroups in terms of a Markov model. We build Markov models from sample data to determine whether human behavior can be described as Markovian and what order model best fits observed data. Constructing mathematical models to understand routine behavior can advance domains ranging from traffic management to healthcare.**

***Keywords*—Pervasive Computing, Smart Home, Routine Behavior Patterns, Markov Model, Human Dynamics**

I. Introduction

To lead healthier and more productive lives, understanding human behavior and its variability across groups is vital. Traditionally, questionnaires and human observations were used to study human behaviors [1]. An insufficient amount of objective, ecologically-valid data limits the ability to develop formal mathematical models to understand routine behaviors.

With our new ability to collect a massive amount of sensor data on subjects in an unobtrusive manner, we can now design data mining methods to better understand human behavior. Researchers have been designing techniques for automatically recognizing activities in complex situations from scripted laboratory conditions to continuous real-world recognition [2], but this work does not offer a mechanistic description of behavioral patterns or insights into variations between populations.

Scientists have been using stochastic models to understand behaviors and identify predictable patterns, including describing a vehicle driver's status [3], studying human social behavior based on cell phone data [4], and modeling occupancy behavior for energy-efficient buildings [5]. But a mechanistic description of stochastic models of routine behavior has not been explored.

Our study provides evidence to support three general principles underlying human behavior in daily environments. First, human routine behavior can be described by formal stochastic processes. Second, data supporting this conclusion can be collected using ambient sensors in everyday environments in an ecologically valid manner. Third, the Markov model can be useful for the study of human dynamics and analyzing differences in behavior among population subgroups.

II. Methodology

With the development of technologies to record a large amount of behavior-based information in everyday settings, researchers now have access to methods for formal modelling activity sequences and timings in an ecologically-valid manner. We collect automatically-recognized activity data from smart home environments [6], [7]. The set of predefined activities that we categorize and use in this analysis are Bathe, Relax, Cook, Eat, Personal Hygiene, Wash Dishes, Sleep, Enter Home, Leave Home, Take Medicine, Bed Toilet Transition (BTT) and Work. We use a separate class, Other Activity, to recognize unidentified sensor events. During the experimental period, change point detection (CPD) is utilized to detect the changes between activities by segmenting the data into sequences that represent single, uninterrupted activities [8].

Routine behaviors can be interpreted using Markov models with conditional probabilities representing activity transitions. In our current study, we focus on the first and second order conditional probabilities. The transition matrix of activities will be calculated by first-order conditional probabilities. The Markov model order will be selected by Akaike information criterion (AIC), Bayesian information criterion (BIC), and efficient determination criterion (EDC) [9]. The AIC method will overestimate the true order, thus we will implement BIC, a consistent method for large samples. We also utilize EDC to select the model order since EDC, as a strong consistent estimate, combines both the AIC and BIC methods.

The goodness of fit (GOF) for the orders selected will be studied by testing two hypotheses: the data represent a partial sample of the process (a sample population) and the data represent a partial sample of some members of the process (population subgroups). A number of GOF test statistics for testing both null hypotheses can be used, including Kolmogorov-Smirnov and Cramér-von Mises test statistics [10].

To evaluate the steps, we first perform the experiments both at an entire population level and among two subgroups: healthy older adults and older adults with chronic cognitive health conditions. We test the hypothesis that differences in health status between subgroups are significantly reflected by patterns of activities. We will investigate hypotheses such as whether the declining of cognitive recognition leads to different Markov model orders.

III. Results

To learn the general principles behind human behavior in everyday environments, we first calculate the first and second order conditional probabilities to explore and understand data. Given three successive activities, we denote the current activity as x_t, the previous activity as x_{t-1}, and the following activity as x_{t+1}. The first order conditional probability is calculated as the number of activities occurring in the order x_t, x_{t+1} divided by the number of all possible values for activity x_{t+1} following x_t(See Equation 1). A similar calculation is performed for the second order conditional probability (Equation 2). Table I shows the transition matrix that based on the first order conditional probabilities from a home with over five years of collected data. Table II is a sample of the second order conditional probabilities from the same resident.

$$P(x_{t+1}|\, x_t) = \frac{the\ number\ of < x_t, x_{t+1} >}{the\ number\ of < x_t, all\ possible\ activities >} \quad (1)$$

$$P(x_{t+1}|\, x_t, x_{t-1}) = \frac{the\ number\ of < x_{t-1}, x_t, x_{t+1} >}{the\ number\ of < x_{t-1}, x_t, all\ possible\ activities >} \quad (2)$$

In Table I, we use bold font for the largest probability in each transition. For example, from the activity Bed Toilent Transition, there is an 86% chance that the next activity is Sleep. This is consistent with our initution: a resident gets up from the bed and uses the toilet. Then, with a large probability, the resident goes back to sleep.

Given the first and second order conditional probabilities, we can calculate the log maximum likelihood of each order and then use it to get the AIC values [9], [11]. For the 1st order Markov model, the AIC value is -22532, and for the 2nd order, the AIC value is -26631. We prefer a model with the lower AIC values. Thus, the Markov model of order two provides a better fit than order one for this smart home. This provides evidence that the resident follows longer sequences of activities in a daily routine.

IV. Discussions

In this study, we construct Markov models for formally modeling human routines based on smart home sensor data. A sample population and population subgroups are used to perform the experiments. We determine whether human behavior can be described as Markovian, select the Markov model order and investigate the goodness of fit for the orders selected. The findings not only can lead to more effective medical interventions, but also may benefit other fields.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 1543656.

References

[1] S. A. Prince, K. B. Adamo, M. E. Hamel, J. Hardt, S. C. Gorber, and M. Tremblay, “A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review,” *Int. J. Behav. Nutr. Phys. Act.*, vol. 5, no. 1, p. 56, 2008.

[2] D. J. Cook, “Activity Learning from Sensor Data.”

[3] S. Di Cairano, D. Bernardini, A. Bemporad, and I. V Kolmanovsky, “Stochastic MPC with learning for driver-predictive vehicle control and its application to HEV energy management.,” *IEEE Trans. Contr. Sys. Techn.*, vol. 22, no. 3, pp. 1018–1031, 2014.

[4] N. Eagle and K. Greene, *Reality mining: Using big data to engineer a better world*. MIT Press, 2014.

[5] K. Sun, D. Yan, T. Hong, and S. Guo, “Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration,” *Build. Environ.*, vol. 79, pp. 1–12, 2014.

[6] N. C. Krishnan and D. J. Cook, “Activity recognition on streaming sensor data,” *Pervasive Mob. Comput.*, vol. 10, no. PART B, pp. 138–154, 2014.

[7] A. Casas, D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan, “Smart Home,” pp. 62–69, 2013.

[8] S. Aminikhanghahi, T. Wang, and D. J. Cook, “Real-time change point detection with application to smart home time series data,” *IEEE Trans. Knowl. Data Eng.*, 2018.

[9] C. C. Dorea, C. R. Goncalves, and P. A. Resende, “Simulation results for Markov model seletion: AIC, BIC and EDC,” in *Proc. World Congress on Engineering and Computer Science*, 2014, vol. 2, pp. 899–901.

[10] M. S. Kaiser, S. N. Lahiri, and D. J. Nordman, “Goodness of fit tests for a class of Markov random field models,” *Ann. Stat.*, vol. 40, no. 1, pp. 104–130, 2012.

[11] R. W. Katz, “On some criteria for estimating the order of a Markov chain,” *Technometrics*, vol. 23, no. 3, pp. 243–249, 1981.

TABLE I. A Transition Matrix

From \ To	Bathe	Bed Toilet Transition	Cook	Enter Home	Eat	Leave Home	Personal Hygiene	Relax	Sleep	Take Medicine	Work	Wash Dishes
Bathe	0.00	0.00	0.05	0.05	0.00	0.10	**0.43**	0.07	0.00	0.00	**0.31**	0.00
Bed Toilet Transition	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	**0.86**	0.00	0.00	0.00
Cook	0.00	0.00	0.04	0.01	0.03	0.01	0.19	0.00	0.00	0.00	**0.62**	0.11
Enter Home	0.00	0.00	0.04	0.01	0.00	0.01	**0.56**	0.02	0.00	0.00	**0.36**	0.01
Eat	0.00	0.00	**0.27**	0.00	0.00	0.02	0.15	0.00	0.00	0.00	**0.48**	0.08
Leave Home	0.00	0.00	0.03	**0.43**	0.00	0.02	**0.34**	0.01	0.00	0.00	0.16	0.00
Personal Hygiene	0.00	0.03	0.13	0.08	0.00	0.05	0.09	0.03	0.10	0.00	**0.48**	0.01
Relax	0.00	0.02	0.05	0.09	0.00	0.06	**0.51**	0.02	0.05	0.00	0.19	0.01
Sleep	0.00	**0.51**	0.01	0.00	0.00	0.00	**0.48**	0.00	0.00	0.00	0.01	0.00
Take Medicine	0.00	0.00	**0.55**	0.00	0.00	0.00	0.09	0.00	0.03	0.00	**0.33**	0.00
Work	0.00	0.00	0.15	0.11	0.00	0.04	**0.53**	0.01	0.01	0.00	0.09	0.06
Wash Dishes	0.00	0.00	**0.26**	0.00	0.01	0.00	0.21	0.00	0.00	0.00	**0.45**	0.06

TABLE II. A Sample Of Second Order Conditional Probabilities

x_{t-1}	x_t	x_{t+1}	$P(x_{t+1}\|\, x_{t-1}, x_t)$
Sleep	BTT	Sleep	0.4285
Work	Relax	Work	0.1893
Bathe	Leave Home	Leave Home	0.5000
Enter Home	Eat	Cook	0.3333
Work	Leave Home	Enter Home	0.2339
Personal Hygiene	Leave Home	Leave Home	0.5087
Relax	Sleep	BTT	0.3429
Wash Dishes	Cook	Cook	0.5125
Cook	Eat	Work	0.2342
Eat	Wash Dishes	Personal Hygiene	0.2895