Adaptive Activity Monitoring with Uncertainty Quantification in Switching Gaussian Process Models

Randy Ardywibowo[1]   **Guang Zhao**[1]   **Zhangyang Wang**[1]   **Bobak Mortazavi**[1]
Shuai Huang[2]   **Xiaoning Qian**[1]
Texas A&M University[1]   University of Washington[2]

Abstract

Emerging wearable sensors have enabled the unprecedented ability to continuously monitor human activities for healthcare purposes. However, with so many ambient sensors collecting different measurements, it becomes important not only to maintain good monitoring accuracy, but also low power consumption to ensure sustainable monitoring. This power-efficient sensing scheme can be achieved by deciding which group of sensors to use at a given time, requiring an accurate characterization of the trade-off between sensor energy usage and the uncertainty in ignoring certain sensor signals while monitoring. To address this challenge in the context of activity monitoring, we have designed an adaptive activity monitoring framework. We first propose a *switching Gaussian process* to model the observed sensor signals emitting from the underlying activity states. To efficiently compute the Gaussian process model likelihood and quantify the context prediction uncertainty, we propose a block circulant embedding technique and use Fast Fourier Transforms (FFT) for inference. By computing the Bayesian loss function tailored to switching Gaussian processes, an adaptive monitoring procedure is developed to select features from available sensors that optimize the trade-off between sensor power consumption and the prediction performance quantified by state prediction entropy. We demonstrate the effectiveness of our framework on the popular benchmark of UCI Human Activity Recognition using Smartphones.

1 INTRODUCTION

Smart health solutions are becoming ever more feasible with the rapid development of sensors and mobile applications that can continuously collect human behavioral data. Indeed, many sensors from various sources such as environmental (Poppe, 2007), body-worn sensors (Lukowicz et al., 2004; Karantonis et al., 2006), or even smartphone based sensors (Anguita et al., 2013) are prevalent for health monitoring applications. With so many ambient sensors collecting different measurements, it becomes important not only to maintain good monitoring accuracy, but also low power consumption, to ensure effective and sustainable monitoring. Such a trade-off between monitoring accuracy and monitoring resource allocation is natural and ubiquitous in many applications (He et al., 2006; Wiser et al., 2008; Wu et al., 2018; Wang et al., 2018).

Such problems require an accurate characterization of the trade-off between monitoring resource usage and monitoring accuracy. In the case of human behavioral and health monitoring using ambient sensors, a power efficient sensing scheme can be achieved by deciding which group of sensors to use at a given time. Here, a trade-off arises between sensor energy usage and the uncertainty in ignoring certain sensor signals while monitoring. To characterize this trade-off, it is necessary to have a model that can quantify the uncertainty of the activity prediction with respect to the sensor measurements over time. With this model, a loss function associated with both energy cost and activity prediction uncertainty can be defined and therefore the adaptive monitoring problem can be solved by a sequential decision process.

Sequential decision process with Markov models have been well studied in the literature. For example, Markov decision process (MDP) and partially observed Markov decision process (POMDP) (Krishnamurthy, 2016) are popular stochastic models for sequential decision process. They can be applied for optimal pre-

ventive maintenance policy (Byon, Ntaimo, and Ding, 2010) and path planning under uncertainty (Morere, Marchant, and Ramos, 2017). Other decision process, for example Bayesian Optimization, have been applied for optimal sensor set selection (Garnett, Osborne, and Roberts, 2010). In this paper, activity detection is based on time series sensor signals, so the sensor selection problem depends on not only the current state, but also the history. The computation of the loss function will therefore be complicated due to the sequential decision nature.

Besides all above challenges in characterizing the resource limitation and monitoring accuracy trade-off, modeling the measurements taken from these ambient sensors is also challenging by itself. For example, smartphone sensor measurements are abundant yet noisy, and thus require efficient computational methods to process effectively. Moreover, the sensor measurement frequency can vary with time, the duration of each activity may vary, and the multivariate effects between the large set of features can be difficult to capture. Methods to model such time-series measurements include linear dynamical systems (Barber, 2012; Ardywibowo et al., 2018), ARMA models (Torres et al., 2005), Kalman filters (Harvey, 1990), point processes (Gunawardana, Meek, and Xu, 2011), and Recurrent Neural Networks (Funahashi and Nakamura, 1993). However, most of these existing methods focused on prediction only and did not attempt to characterize the uncertainty of the measurements.

To tackle this problem, we derive a switching multivariate Gaussian process model for the goal of activity recognition. Our model is a hierarchical one consisting of a Hidden semi-Markov model (HSMM) for the discrete activity states, and a multivariate Gaussian process to model both the time dependent and the inter-variable correlations between different sensor measurements. We use a block circulant embedding technique and use Fast Fourier Transforms (FFT) to speed up model inference and uncertainty quantification of our proposed model. Using this model, we then develop an adaptive monitoring scheme that for each monitoring period uses the optimal group of sensors by optimizing the Bayesian cost function considering both the activity predictive entropy[1] and the energy cost of selected sensors.

We implement our model on the UCI Human Activity Recognition using Smartphones dataset (Anguita et al., 2013). This dataset consists of labeled trajec-

---

[1]Here the predictive entropy is a measure for uncertainty, the reduction of which indicates the information provides by the selected sensors. It is computed approximately by Monte Carlo since it has no closed-form expression.

tories of smartphone sensor measurements from multiple subjects under 6 different activities: walking, walking upstairs, walking downstairs, sitting, standing, and laying. The time series trajectories consist of features extracted from gyroscope and accelerometer measurements, such as movement angle, jerk, acceleration, and moving averages. Extensive results demonstrate the effectiveness of our framework on achieving competitive performance-energy trade-offs

2 THE MODEL: SWITCHING GAUSSIAN PROCESS

Throughout our presentation, we will use the following notation convention: bold faces indicate vectors or multivariate processes, capital letters indicate matrices or covariance functions of two variables, and script letters indicate sets.

2.1 Model Formulation

We first describe the model for time series sensor measurements from one subject. For the different activity types, we assume that there is an underlying semi-Markov jump process (Yu, 2010) that governs the transitions between them. Denoting the discrete valued activity state at time t as $x(t)$, the semi-Markov jump process can be expressed as:

$$x(t) = \sum_{n=1}^{N} x_n \mathbf{1}_{\lambda_n \leq t < \lambda_{n+1}} \qquad (1)$$

Here, $x_n \in \{1, \dots A\} \equiv \mathcal{X}$ are discrete activity states that discriminate between the A different activity types, while λ_n and λ_{n+1} are the random start and end time points of the discrete activity state X_n. We can denote the random duration that the subject stays in X_n as $s_n = \lambda_{n+1} - \lambda_n$. We adopt an explicit-duration model for the sojourn times by modeling s_n using a Gamma distribution with state specific parameters as follows:

$$s_n \mid (x_n = i) \sim Gamma(\theta_i) \quad i \in \mathcal{X} \qquad (2)$$

where $\theta_i = \{k_i, \tau_i\}$ are the parameters for the duration distribution of activity i with k_i and τ_i representing shape and scale parameters respectively. The advantage of this semi-Markov model is that it explicitly models the varying durations of different activity types. For example, in our application, we can see that the walking upstairs and walking downstairs activities all take less time to complete compared to the other activities. This activity duration may be informative in inferring the different activities. To complete the semi-Markov jump process model modeling the underlying activity states, the state transition probability

themselves can be modeled by a transition probability matrix as follows:

$$P(x_{n+1} = j | x_n = i) = p_{ij}. \tag{3}$$

For the observation process $\mathbf{y}(t) = [y^1(t), ..., y^P(t)]^\top$, we incorporate a switching P-variate Gaussian Process. This process models the dynamic heterogeneity of the time series measurements by switching between different Gaussian process models for each inferred activity state or context. Specifically, we model $\mathbf{y}_n(t)$, the observation process occurring at time period $[\tau_{n+1}, \tau_n]$, as follows:

$$\mathbf{y}_n(t) | (x_n = i) \sim \mathcal{GP}(\mathbf{m}_i(t), \mathbf{K}_i(t, t')). \tag{4}$$

Here, conditioned on the activity state $x_n = i$, the observation process $\mathbf{y}(t)$ will be a Gaussian process with activity-state-dependent mean and covariance function $\mathbf{m}_i(t)$ and $\mathbf{K}_i(t, t')$ respectively. The covariance function $\mathbf{K}_i(t, t')$ is a multivariate covariance function between two univariate variables $y_n^a(t)$ and $y_n^b(t)$ in $\mathbf{y}_n(t)$ for a, b in the set of features $\mathcal{Y} = \{1, \ldots, P\}$.

With this, the observation process can be expressed as

$$\mathbf{y}(t) = \sum_{n=1}^{N} \mathbf{y}_n(t) \mathbf{1}_{\{\tau_n \le t < \tau_{n+1}\}}. \tag{5}$$

In order to handle the potential multivariate correlations between the sensor observations, we implement an intrinsic correlation model (Bonilla, Chai, and Williams, 2008). In this model, we assume that we can separate the covariance function of $\mathbf{y}_n(t)$ into a time dependent covariance function $K_i^{\mathcal{T}}(t, t')$, and a free-form inter-variable covariance matrix $K_i^{\mathcal{Y}}(a, b)$ between two variables $y_n^a(t)$ and $y_n^b(t)$ in $\mathbf{y}_n(t)$ as follows:

$$K_i(t, t') = K_i^{\mathcal{T}}(t, t') \otimes K_i^{\mathcal{Y}}(a, b). \tag{6}$$

Knowing the activity state and a set of noisy observations z at a set of time points S, the prediction on a new set of points S^p for the lth variable is given by:

$$y_n^l(S^p) | (\mathbf{z}, x_n = i) \sim N(m_i^l(S^p), \Sigma_i^l(S^p)); \tag{7}$$

$$m_i^l(S^p) = (\mathbf{k}_i^{l,\mathcal{Y}} \otimes K_i^{(S,S^p),\mathcal{T}})^\intercal \Sigma_i^{-1} \mathbf{z}; \tag{8}$$

$$\Sigma_i = K_i^{\mathcal{Y}} \otimes K_i^{\mathcal{T}} + D \otimes I; \tag{9}$$

$$\Sigma_i^l(t_p) = \mathbf{k}_i^{l,\mathcal{Y}} \otimes K_i^{(S,S),T} - (\mathbf{k}_i^{l,\mathcal{Y}} \otimes K_i^{(S,S^p),\mathcal{T}})^\intercal \times \\ \Sigma_i^{-1}(\mathbf{k}_i^{l,\mathcal{Y}} \otimes K_i^{(S,S^p),\mathcal{T}}). \tag{10}$$

Here, $\mathbf{z}$ is a vector of noisy multivariate measurements structured as $\mathbf{z} = [z_1^1, z_2^1, ..., z_N^1, z_1^2, ..., z_N^P]^T$, where z_t^p is the measurement of variable p at time t.
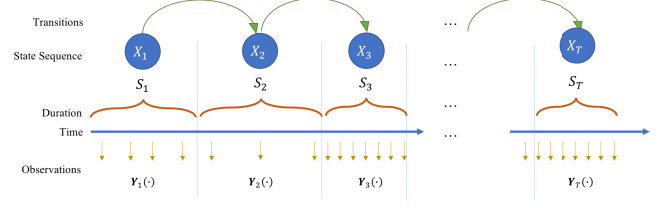

Figure 1: The hidden semi-Markov model.

The vector $\mathbf{k}_i^{l,\mathcal{Y}}$ selects the lth column of $K_i^{\mathcal{Y}}$, while $K_i^{(S,S^p),\mathcal{T}}$ is a matrix of time dependent covariances between the prediction time points and the observed time points. Finally, D is a $P \times P$ matrix of independent noise variances σ_p^2 for each multivariate variable p. The parameters of our model can be combined as: $\boldsymbol{\theta} = \{\boldsymbol{\gamma}_i, p_{ij}, K_i^{\mathcal{T}}(\cdot), K_i^{\mathcal{Y}}(\cdot), \forall, j \in \mathcal{X}\}$. Where the covariance matrices can be further parameterized. Specifically, we use a Matern covariance function for the time dependent covariance, while a Cholesky factorization was used for the multivariable covariance to ensure positive definiteness. An illustration summarizing the model formulation is shown in Figure 1.

2.2 Model Inference

We describe both the parameter inference and process inference algorithms for the proposed model with switching Gaussian process. For parameter inference, we use a population model to combine the data from multiple subjects, treating each subject as an independent realization of our model. We estimate the parameters of our model using maximum likelihood inference. Meanwhile, process inference can be efficiently done, similarly to classical filtering methods in traditional Hidden Markov Models (HMMs) (Rabiner, 1986) using the Forward-Backward algorithm.

2.2.1 Parameter Inference

In our application, since all of the activity states are labeled in the time-series data, estimation of the semi-Markov jump process parameters is straightforward. Specifically, for each Gamma distributed activity state duration with parameter $\boldsymbol{\gamma}_i = \{k_i, \beta_i\}$, we can derive the corresponding maximum likelihood estimates for k_i and β_i as follows:

$$v_i = \log(\frac{1}{N_i} \sum_{n=1}^{N_i} s_i^n) - \frac{1}{N_i} \sum_{n=1}^{N_i} s_i^n; \tag{11}$$

$$\hat{k}_i \approx \frac{3 - v_i + \sqrt{(v_i - 3)^2 + 24v_i}}{12v_i}; \tag{12}$$

$$\hat{\beta}_i = \frac{1}{\hat{k}_i N_i} \sum_{n=1}^{N_i} s_i^n, \tag{13}$$

where s_i^n is the duration of the n-th time stamp in state i. Besides, the transition probability can be inferred by simply counting the number of transitions between activity states as follows:

$$p_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \tag{14}$$

where n_{ij} is the number of transitions from state i to state j.

To ensure positive definiteness of the multivariate covariance, we can parametrize it using the Cholesky decomposition $K_i^{\mathcal{Y}} = LL^\top$ for lower-triangular matrix L. Since each subject's process model is independent from each other given the model parameters, we can write the complete log-likelihood of all subjects as a sum of individual log-likelihoods. Specifically, denote by $\mathcal{Q}_j$ as two times the negative log-likelihood of subject j. This term can be expanded as follows: First, for each time frame λ corresponding to a different activity state, we put our observations vector z in an $N \times P$ matrix form, denoted by Z_λ. Then, denote the matrix of corresponding mean functions for each Z_λ as F_λ.

$$\mathcal{Q}_j = \sum_{\lambda=1}^{T_j} N \log |K_{i_\tau}^Y| + P \log |K_{i_\tau}^T| + N P \log 2\pi$$

$$tr[(K_{i_\tau}^Y)^{-1} F_\lambda (K_{i_\tau}^T)] F_\lambda^\top] + N \sum_{p=1}^{P} \log \sigma_p^2 +$$

$$tr[(Z_\lambda - F_\lambda) D^{-1} (Z_\lambda - F_\lambda)^\top] \tag{15}$$

where i_τ is the activity state of subject j in time frame λ with corresponding time dependent and multivariate covariance matrices $K_{(i_\tau)}^{\mathcal{T}}$ and $K_{(i_\tau)}^{\mathcal{Y}}$. With this, the complete log-likelihood $\mathcal{Q}$ will be a sum of the individual log-likelihoods:

$$2\mathcal{Q} = \sum_{j=1}^{M} \mathcal{Q}_j \tag{16}$$

With this, any optimization method can be used to estimate the parameters of the multivariate Gaussian process model. The exact parameter updates are omitted from this presentation and the reader is referred to (Bonilla, Chai, and Williams, 2008) for more details.

Fast Inference using Block Circulant Embedding Computing the likelihood as well as the gradient of the multivariate Gaussian Process models can be computationally challenging when dealing with a large number of features and the potential interactions between them. This is mainly due to the inversion and determinant calculation operations that need to be performed on the covariance matrix $\mathbf{K}_i(t, t')$.

To speed up these computations, we propose a block circulant embedding approach. We first note that the measurements of our dataset are evenly spaced in time and assume that we only have one feature measurement. For this setting, the covariance matrix can be expressed as follows:

$$C = \begin{matrix} C(0) & C(1) & C(2) & & C(T) \\ C(1) & C(0) & C(1) & & C(T-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C(T) & C(T-1) & C(T-2) & & C(0) \end{matrix}$$

Here $C(t)$ are scalars as we only have a single feature. Observe that we can embed this matrix in the following larger circulant matrix: $C =$

$$\begin{bmatrix} C(0) & \cdots & C(T) & C(T-1) & \cdots & C(1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C(T) & \cdots & C(0) & C(1) & \cdots & C(T-1) \\ C(T-1) & \cdots & C(1) & C(0) & \cdots & C(T-2) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C(1) & \cdots & C(T-1) & C(T-2) & \cdots & C(0) \end{bmatrix},$$

which is fully specified by the first row vector $\mathbf{c} = [C(0) \; C(1) \; \cdots \; C(T) \; C(T-1) \; \cdots \; C(1)]$. We can perform an eigenvalue decomposition of the above matrix as follows:

$$C = F \Lambda F^* \tag{17}$$

F and F^* denote the Fourier and inverse Fourier transform respectively, and $\Lambda = F \mathbf{c}$. Using the Fast Fourier Transform (FFT), this calculation can be done in $O(T \log(T))$. Consequently, inverse and determinant calculations can be done at this same time complexity (Davis, 2012). This idea can be extended to the case of block circulant matrices. First we note that the covariance matrix $K_i(t, t')$ can be embedded into a block circulant matrix specified by the matrix $\mathbf{K}_i = [K_i^{\mathcal{T}}(0) \; \cdots \; K_i^{\mathcal{T}}(T) \; K_i^{\mathcal{T}}(T-1) \; \cdots \; K_i^{\mathcal{T}}(1)] \cdot K_i^{\mathcal{Y}}$. We then notice that each vector $[K_i^{\mathcal{T}}(0) \; \cdots \; K_i^{\mathcal{T}}(T) \; K_i^{\mathcal{T}}(T-1) \; \cdots \; K_i^{\mathcal{T}}(1)] K_i^{\mathcal{Y}}(a, b) \; a, b \in 1 \cdots P$ defines a circulant submatrix. To compute the determinant and eigen-decomposition of each block circulant matrix, it suffices to simply decompose each of them using FFT. Finally, the model likelihood can be computed in $O(P^2 T \log(T))$.

2.2.2 Process Inference

Activity state inference can be conducted using standard filtering approaches similar to classical Hidden Markov Models (HMM) (Rabiner, 1986). Specifically, we are interested in estimating the current state at time t given all previous observations.

Firstly, denote $y(t_0 : t_1)$ as a collection of variables y_λ for $t_0 \leq \lambda \leq t_1$. The expression $x[t_1 : t_2] = i$ indicates
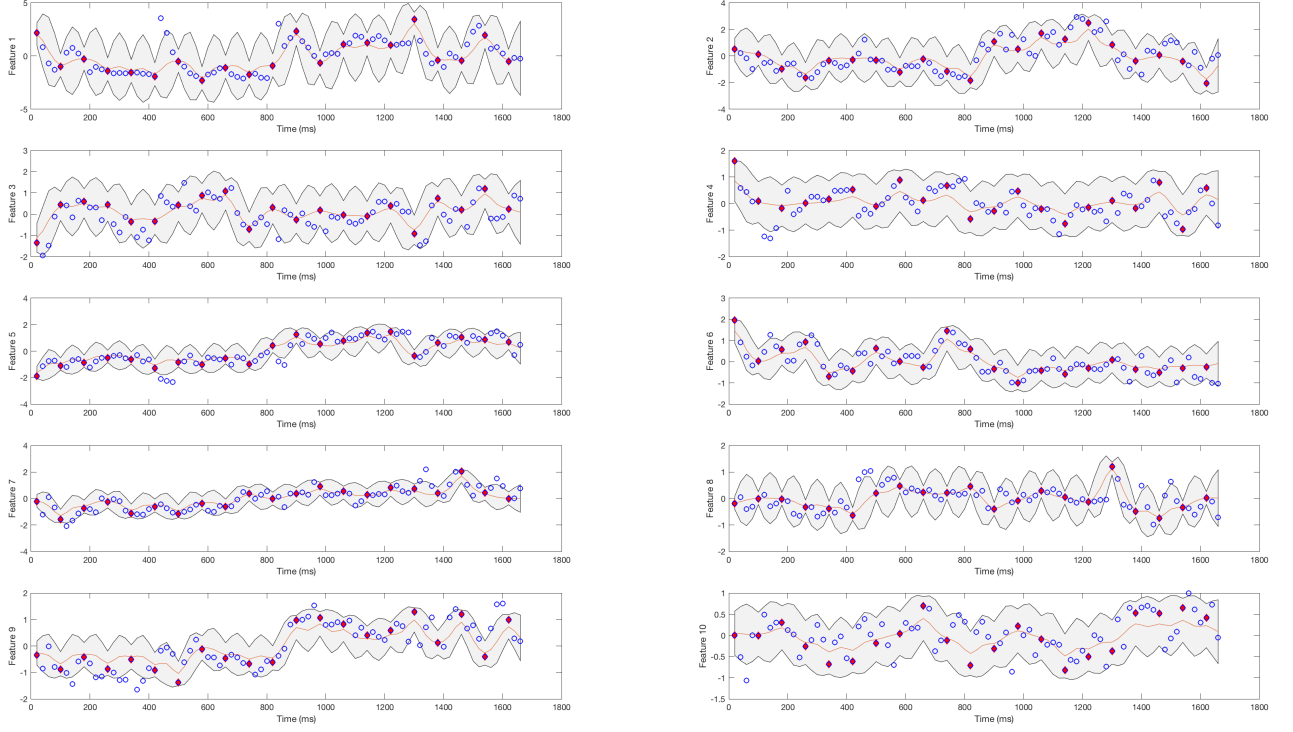
Figure 2: Prediction trajectory comparison of different features.

that x is in state i between times t_1 and t_2 but no later or sooner. In other words, x will be in a different state before and after time t_1 and t_2 respectively. With this, the activity detection problem can be expressed by the following maximum *a posteriori* (MAP) filtering problem:

$$x_t = argmax_{i \in \mathcal{X}} P(x_t = i | \mathbf{y}_{t_0:t}) \qquad (18)$$

It can be solved recursively using dynamic programming. Specifically, at any given time point we need to compute the following forward and backward variables respectively:

$$\alpha_t(j, d) := P(x_{[t-d+1:t]} = j | \mathbf{y}_{1:t}); \qquad (19)$$

$$\beta_t(j, d) := P(\mathbf{y}_{t+1:T} | x_{[t-d+1:t]} = j) \qquad (20)$$

Since we are merely interested in filtering the current activity states or contexts in an online fashion, we need only compute the forward variables $\alpha_t(j, d)$. The recursion for computing them can be expressed as follows:

$$\alpha_t(j, d) \qquad (21)$$
$$= \sum_{i \in \mathcal{X}} \sum_{j} \sum_{d' \in \mathcal{D}} \alpha_{t-d}(j, d') a_{(i, d')(j, d)} b_j(\mathbf{y}_{t-d+1:t})$$

Here, $a_{(i, d')(j, d)}$ is the transition probability from staying in state i for duration, d', towards staying in state

j for duration d. Based on our model, this is a semi-Markov jump process, with state specific durations modeled as gamma distributions. Hence, this probability can be expressed as:

$$a_{(i, d')(j, d)} := P(x_{[t-d+1:t]} = j | x_{[t-d-d'+1:t-d]} = i)$$
$$= Gamma(d | \theta_i) P(x_{t-d+1} = j | x_{t-d} = j)$$
$$Gamma(d | \theta_j)$$
$$= Gamma(d | \theta_j) P_{ij} Gamma(d | \theta_j) \qquad (22)$$

In (21), $b_{(j, d)}(\mathbf{y}_{(t-d+1:t)})$ denotes the observation probability that is modeled as a switching Gaussian process. Specifically, this quantity can be expressed as follows:

$$b_j(\mathbf{y}_{t+1:t+d}) := P(\mathbf{y}_{t+1:t+d} | x_{[t+1:t+d]} = j) \qquad (23)$$
$$= N(\mathbf{y}_{t+1:t+d} | \mathbf{m}_j^{t+1:t+d}, K_j^{t+1:t+d})$$

where $\mathbf{m}_j^{(t+1:t+d)}$ and $K_j^{(t+1:t+d)}$ are the predicted means and covariances from time points $t+1$ to $t+d$.

3   THE ADAPTIVE MONITORING FRAMEWORK

In this section, an adaptive monitoring method is proposed based on the hidden semi-Markov model with switching Gaussian Process. In particular, with the previous process inference procedure, we can derive

the activity state predictive distribution. The prediction uncertainty can then be characterized by the entropy of the derived distribution. In the cases when we are confident about the activity state in the next time point based on the predictive entropy, we may not need all the sensor data for the next observations, and therefore we can choose only a subset of the sensors for the sake of saving energy. Adaptive monitoring will determine which group of sensor measurements may not be necessary, if we predefine groups of features based on the senor monitoring resource allocation requirements.

The set of predefined feature groups can be denoted as $\mathcal{F}$, from which one of the feature group $m \in \mathcal{F}$ is chosen at each observation time t to observe signals $\mathbf{y}_t^m$. For each feature group m, we also define its cost related to power consumption, which is set to a constant ℓ_m for simplicity. The selected feature group should balance the energy cost and prediction uncertainty gain. This is achieved by minimizing the following loss over all predefined feature groups:

$$\mathcal{L}(m) = \mathbb{E}_{P(\mathbf{y}_{t+1}^m \mid \mathbf{y}_{1:t}^{\mathcal{F}})}[H[P(x_{t+1} \mid \mathbf{y}_{1:t}^{\mathcal{F}}, \mathbf{y}_{t+1}^m)]] + \ell_m \quad (24)$$

where $\mathbf{y}_{1:t}^{\mathcal{F}}$ denotes all the previous observations of feature groups selected from $\mathcal{F}$ following the same rule, $H[p(x)] = -\int p(x)\log p(x)$ is the entropy function. In the loss function, the first term averages over the predictive distribution of $\mathbf{y}_{t+1}^m$ given $\boldsymbol{y}_{1:t}^{\mathcal{F}}$ and it describes the remaining uncertainty of x_{t+1} after observing $\mathbf{y}_{t+1}^m$, we would like to minimize this value to make sure that ignoring feature group m does not increase much prediction uncertainty. At the same time, we also would like to achieve small energy cost ℓ_m.

Loss Function Computation The minimization of (24) does not have an analytic solution form and must be solved approximately. We first approximate the expected predictive entropy in (24) by drawing samples $\boldsymbol{y}_{t+1}^{m(i)}$ from $P(\mathbf{y}_{t+1}^m \mid \mathbf{y}_{1:t}^{\mathcal{F}})$. For each of these samples, we then calculate the corresponding entropy function $H[P(x_{t+1} \mid \mathbf{y}_{1:t}^{\mathcal{F}}, \mathbf{y}_{t+1}^{m(i)})]$.

To derive the sampling probability $P(\mathbf{y}_{t+1}^m \mid \mathbf{y}_{1:t}^{\mathcal{F}})$, we adopt the following message passing algorithm:

$$P(\mathbf{y}_{t+1}^m \mid \mathbf{y}_{1:t}^{\mathcal{F}}) \quad (25)$$
$$= \sum_{i,j} \sum_{\mathcal{X}} \sum_{d',d \in D} b_j(\mathbf{y}_{t+1}^m) a_{(i\,d')(j\,d)} \gamma_{t\,d+1}(j\,d)$$

which gives a Gaussian mixture distribution with respect to $\mathbf{y}_{t+1}^m$. For each samples $\boldsymbol{y}_{t+1}^{m(i)}$, we can easily calculate $P(\mathbf{y}_{t+1}^m \mid \mathbf{y}_{1:t}^{\mathcal{F}}, \mathbf{y}_{t+1}^{m(i)})$ based on (19). Then the loss function can be approximated by:

$$\mathcal{L}(m) = \ell_m + \frac{1}{N}\sum_{i=1}^{N} H[P(x_{t+1} \mid \mathbf{y}_{1:t}^{\mathcal{F}}, \mathbf{y}_{t+1}^{m(i)})] \quad (26)$$

where N is the number of samples drawn from $P(\mathbf{y}_{t+1}^m \mid \mathbf{y}_{1:t}^{\mathcal{F}})$.

4 RESULTS AND DISCUSSION

We apply our model to the popular UCI benchmark of Human Activity Recognition using Smartphones (Anguita et al., 2013). This dataset contains sensor measurements from a group of 30 volunteers performing six types of activities of daily living, including walking (1), walking upstairs (2), walking downstairs (3), sitting (4), standing (5), and laying (6). The sensor measurements are all taken at a constant rate and labeled manually with the corresponding activity type. We pre-process the data by Principal Component Analysis (PCA) to reduce computation overhead. Specifically, 10 principal components are derived from the 561 original features. We assume a constant mean for each activity by computing the population mean corresponding of each.

We evaluate the performance through two sets of evaluations. For the first evaluation, we assume activity states are known, and evaluate by the mean squared error (MSE) and absolute difference error (ABS) for trajectory prediction. In our second evaluation, we conduct joint activity recognition and trajectory prediction, where we only have observations up to the time of prediction. Several different Gaussian process setups and assumptions are compared to select the best performing model.

Table 1: Prediction result comparison at different Gaussian process setups.

Model Setup	MSE	ABS
Baseline + Separate time dependence + Separate Multivariate	0.4988	0.4480
Baseline + Separate time dependence + Combined Multivariate	0.4040	0.4244
Separate time dependence + Combined Multivariate	**0.3852**	**0.4235**

4.1 Trajectory Performance Prediction assuming Known Activity States

We compare the sensor signal trajectory prediction performance under three different Gaussian process model assumptions. The first assumption uses two Gaussian process components: a baseline model to model the entire trajectory of all patients and an activity specific model to capture the varying dynamics of the different activities. We assume that either the temporal and multivariate dependencies are modeled separately for each activity, or the multivariate dependencies are combined between different activities. The
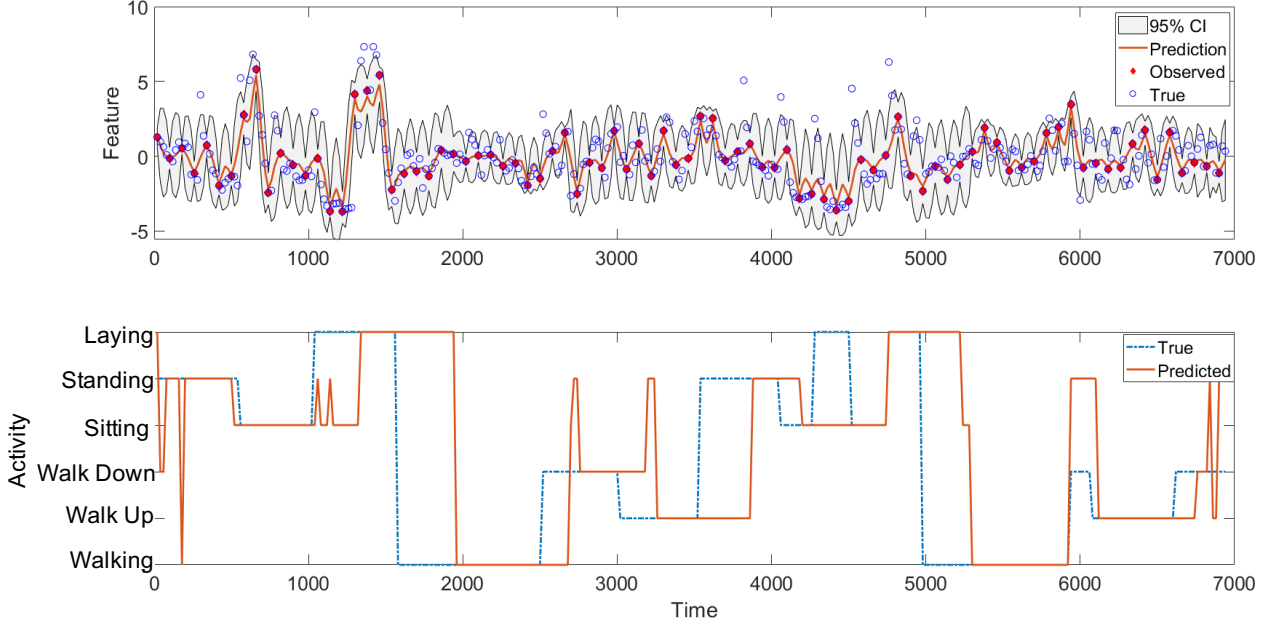
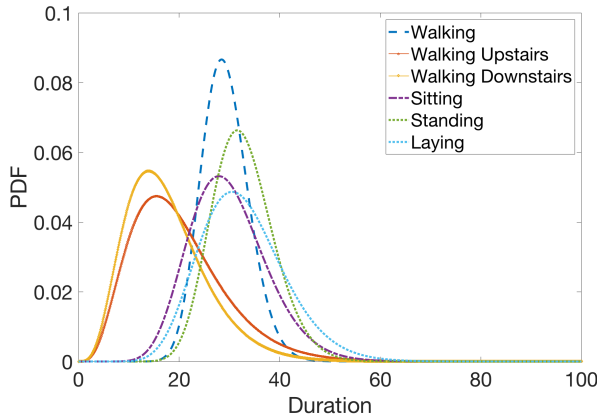Figure 3: Prediction trajectory for activity recognition and prediction.

Figure 4: Learned Gamma duration distributions of the different activity types

Figure 5: Energy cost and accuracy trade-off curve for activity recognition

prediction accuracy of these two models are shown in the first two entries of Table 1. Our best performing model, however, considers only a single multivariate effect model for different activities with no special consideration for any baseline trajectory. We use this best performer's assumption for the remaining experiments. We set the ratio between observed measurements and held-out measurements to be 1:4. Figure 2 demonstrates prediction results of different variables where the model is able to follow the trajectories of the held-out data accurately. Moreover, the sensor signals for different activity states also have different variances, indicated by varying widths of the predicted confidence intervals.
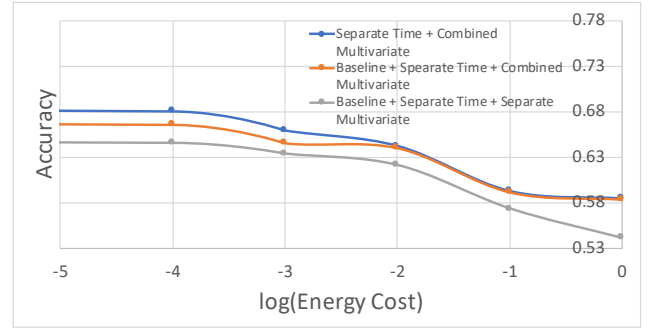
We compare the activity state prediction results in two tasks. Figure 4 illustrates the first task, where the learned semi-Markov jump process duration distributions manifest a difference in duration between walking upstairs and downstairs compared to the other activities. Figure 3 demonstrates activity state prediction results in our second task. In this task, the average testing prediction accuracy can go up to 74.21%. Additionally, a large portion of inaccurate estimations are due to the time lag between the actual and predicted activity switch, which is a common problem in filtering tasks that need a certain amount of measurements in order to be confident in an activity switch.

4.2 Adaptive Monitoring
Finally, We implement the adaptive monitoring on the same UCI dataset with same training and test-
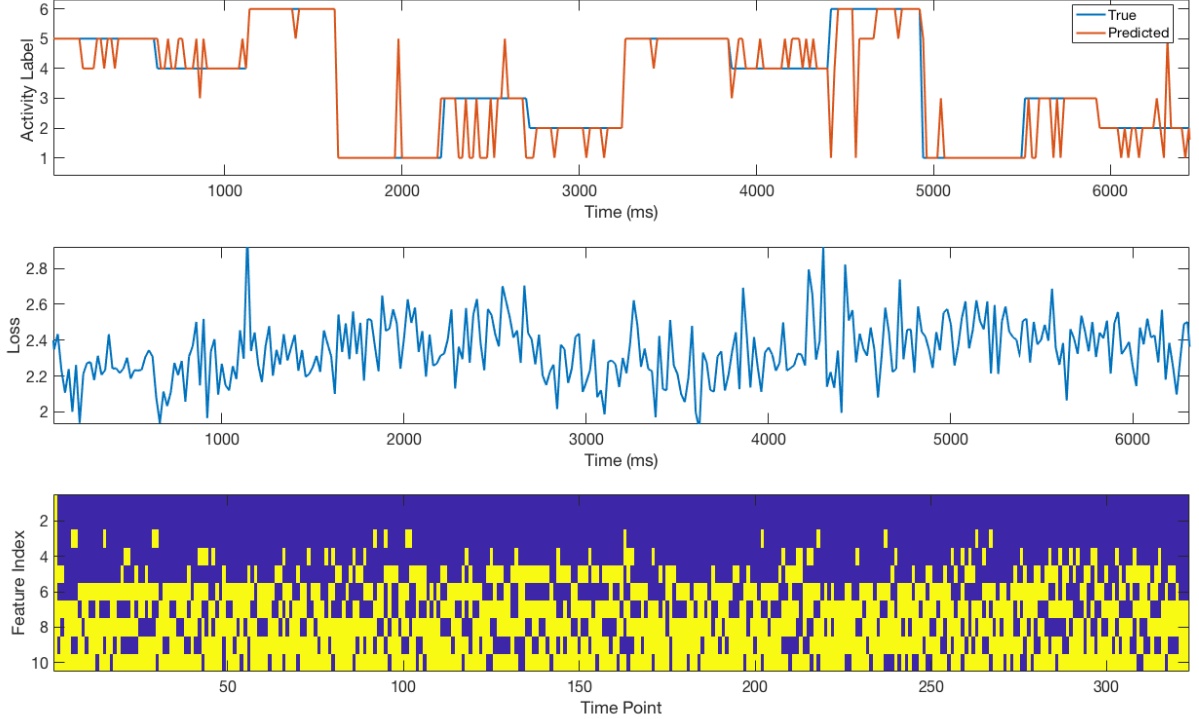
Figure 6: Context prediction results, loss function values, and active features across time for $\lambda = 0.1$. Yellow highlights indicate active sensors.

ing setup. To perform sensor selection, we treat the 10 principal components as 10 different sensors and assume they have the same energy cost. We then define feature subsets consisting of all possible combinations of 4, 7, and 10 sensor measurements. Thus, the energy cost for each group is proportional to the number of measurements in the group. In our experiments, we vary the energy cost λ of a single sensor from 0 to 1 and view its impact on prediction performance.

As shown in Figure 5, we see that as the energy cost λ is increased, the prediction accuracy decreases as each sensor measurement comes at a higher cost. On the other hand, when there is no energy cost, the monitoring plan will prefer to utilize all of the sensor measurements available to perform the best prediction possible. A sample of the context prediction trajectory with energy cost $\lambda = 0.1$, which exhibits the most activity in adaptive feature selection, is shown in Figure 6. In this specific example, we are able to get an accuracy of 79.26% with an average sensor usage of 73.42%. The most used features are the leading principal components indexed starting from 10. This indicates that the leading principal components are better correlated to the activity recognition task compared to lower principal components. Moreover, this also shows that our adaptive monitoring framework can detect the most relevant principal component features without knowing their order in advance.

5 CONCLUSION

In this paper, we propose a hierarchical model consisting of a multivariate switching Gaussian process to model the signals based on different activity types. We applied our model on trajectory and activity prediction with the UCI dataset for model verification. MSE for trajectory prediction can be as small as 0.3852, and activity recognition accuracy can reach 74.21%. Based on this model, we proposed an adaptive monitoring approach balancing the trade-off between sensor energy cost and prediction uncertainty. Within this monitoring scheme, we characterize the trade-off between monitoring accuracy and sensor energy efficiency. We show that our activity recognition scheme can stay robust and perform well under energy restrictions.

References

Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; and Reyes-Ortiz, J. L. 2013. A public domain dataset

for human activity recognition using smartphones. In *ESANN*.

Ardywibowo, R.; Huang, S.; Gui, S.; Xiao, C.; Cheng, Y.; Liu, J.; and Qian, X. 2018. Switching-state dynamical modeling of daily behavioral data. *Journal of Healthcare Informatics Research* 2(3):228 247.

Barber, D. 2012. *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

Bonilla, E. V.; Chai, K. M.; and Williams, C. 2008. Multi-task gaussian process prediction. In *Advances in neural information processing systems*, 153 160.

Byon, E.; Ntaimo, L.; and Ding, Y. 2010. Optimal maintenance strategies for wind turbine systems under stochastic weather conditions. *IEEE Transactions on Reliability* 59(2):393 404.

Davis, P. J. 2012. *Circulant matrices*. American Mathematical Soc.

Funahashi, K.-i., and Nakamura, Y. 1993. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks* 6(6):801 806.

Garnett, R.; Osborne, M. A.; and Roberts, S. J. 2010. Bayesian optimization for sensor set selection. In *Proceedings of the 9th ACM/IEEE international conference on information processing in sensor networks*, 209 219. ACM.

Gunawardana, A.; Meek, C.; and Xu, P. 2011. A model for temporal dependencies in event streams. In *Advances in Neural Information Processing Systems*, 1962 1970.

Harvey, A. C. 1990. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

He, T.; Krishnamurthy, S.; Luo, L.; Yan, T.; Gu, L.; Stoleru, R.; Zhou, G.; Cao, Q.; Vicaire, P.; Stankovic, J. A.; et al. 2006. Vigilnet: An integrated sensor network system for energy-e cient surveillance. *ACM Transactions on Sensor Networks (TOSN)* 2(1):1 38.

Karantonis, D. M.; Narayanan, M. R.; Mathie, M.; Lovell, N. H.; and Celler, B. G. 2006. Implementation of a real-time human movement classi er using a triaxial accelerometer for ambulatory monitoring. *IEEE transactions on information technology in biomedicine* 10(1):156 167.

Krishnamurthy, V. 2016. *Partially Observed Markov Decision Processes*. Cambridge University Press.

Lukowicz, P.; Ward, J. A.; Junker, H.; Stager, M.; Troster, G.; Atrash, A.; and Starner, T. 2004. Recognizing workshop activity using body worn microphones and accelerometers. In *International conference on pervasive computing*, 18 32. Springer.

Morere, P.; Marchant, R.; and Ramos, F. 2017. Sequential bayesian optimization as a pomdp for environment monitoring with uavs. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, 6381 6388. IEEE.

Poppe, R. 2007. Vision-based human motion analysis: An overview. *Computer vision and image understanding* 108(1-2):4 18.

Rabiner, L. R. 1986. An introduction to hidden markov models. *ieee assp magazine* 3(1):4 16.

Torres, J. L.; Garcia, A.; De Blas, M.; and De Francisco, A. 2005. Forecast of hourly average wind speed with arma models in navarre (spain). *Solar Energy* 79(1):65 77.

Wang, Y.; Nguyen, T.; Zhao, Y.; Wang, Z.; Lin, Y.; and Baraniuk, R. 2018. Energynet: Energy-e cient dynamic inference. *NIPS Workshop on Compact Deep Neural Networks with Industrial Applications*.

Wiser, R.; Bollinger, M.; Barbose, G.; Belyeu, K.; Hand, M.; Heimiller, D.; Lew, D.; Milligan, M.; Mills, A.; Moreno, A.; et al. 2008. Annual report on us wind power installation, cost, and performance trends: 2006.

Wu, J.; Wang, Y.; Wu, Z.; Wang, Z.; Veeraraghavan, A.; and Lin, Y. 2018. Deep k-means: Re-training and parameter sharing with harder cluster assignments for compressing deep convolutions. In *International Conference on Machine Learning*, 5359 5368.

Yu, S.-Z. 2010. Hidden semi-markov models. *Artificial intelligence* 174(2):215 243.