Automated health monitoring using AI

Kiranmoy Das

Beijing Institute of Mathematical Sciences and Applications

July 9, 2024

Outline

Outline

Outline

Outline

Outline

Motivation

- “The Internet of things (IoT) describes the network of physical objects that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the Internet."
- A number of recent research works illustrate the usefulness and efficiency of IoT applications to increase the quality of health services to the citizens of smart cities.
- Wearable biosensors are hugely used in monitoring patients with chronic diseases. This type of system can monitor patients’ health conditions not only in hospitals and/or medical centers but in their own personal environments as well.

Importance of IoT

- Remote monitoring of patients are advantageous since it reduces patients' discomfort and risk of infection due to long stay in hospitals, and also offers mobility. This is very relevant in the post-covid period.
- For the developing countries where most of the doctors reside in the urban areas, digitised healthcare services through biosensors are promoted especially for the patients living in the rural areas.
- Biosensors are wearable as well as implantable, and they provide measurements on basic important physiological parameters (hereafter referred to as biomarkers) e.g. heart rate, blood pressure, body and skin temperature, oxygen saturation, respiration rate etc., as well as environmental parameters e.g. location, temperature, humidity, light etc.

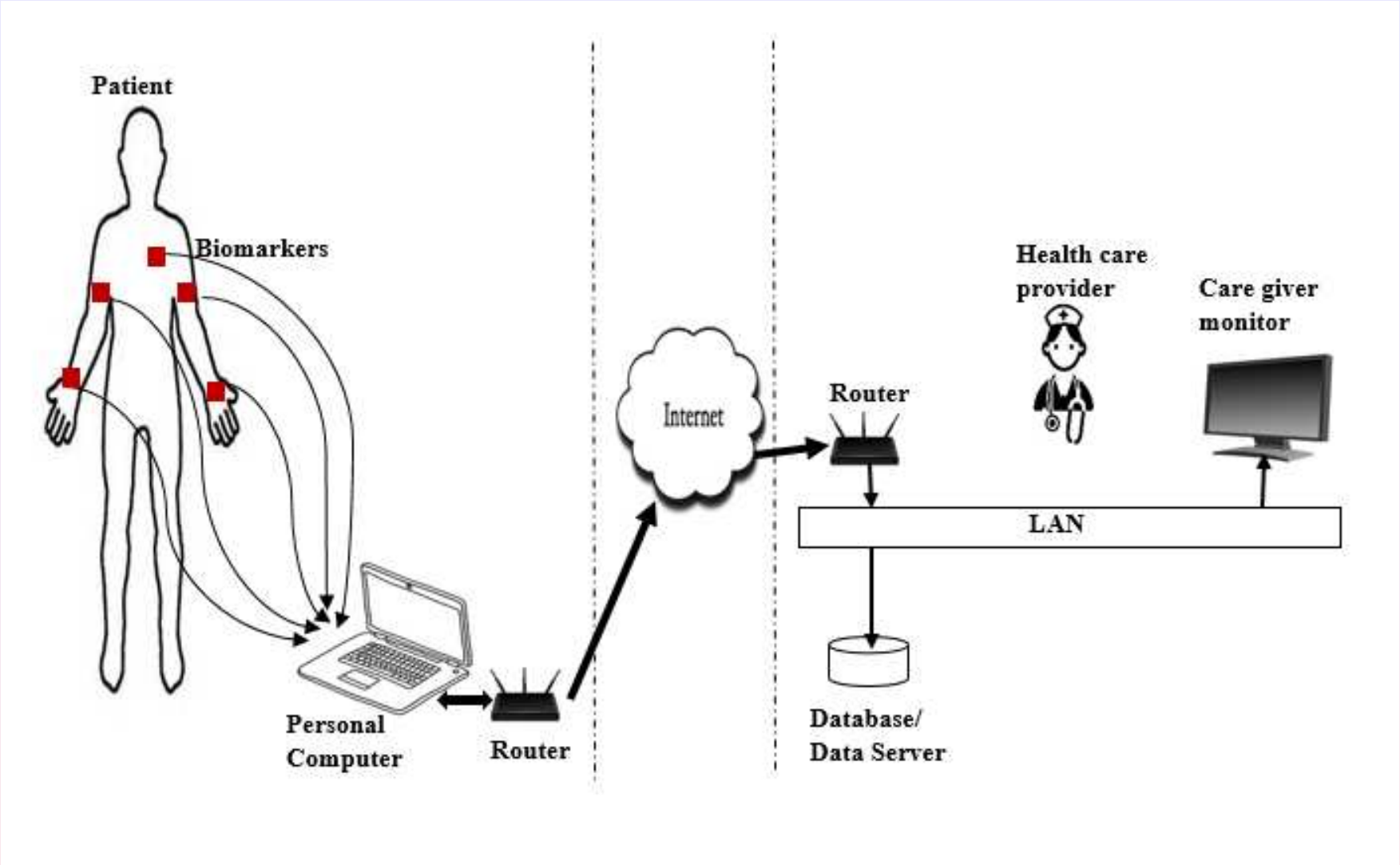

Figure: Health monitoring using AI

Our Goal

- We propose a statistical model that can predict a patient's health condition in the next time point based on the previously observed measurements.
- We consider a set of correlated biomarkers which are measured longitudinally from a patient's body and are recorded automatically by a set of sensor nodes placed at different parts of the patient's body.
- These sensor nodes, which are low-powered tiny devices, convert the continuous measurements to binary or ordinal outcomes (say; low, good, fair, high, very high) based on some (known) prefixed thresholds and send to the base station.
- The challenge is to use these ordinal outcomes for an effective patient monitoring.

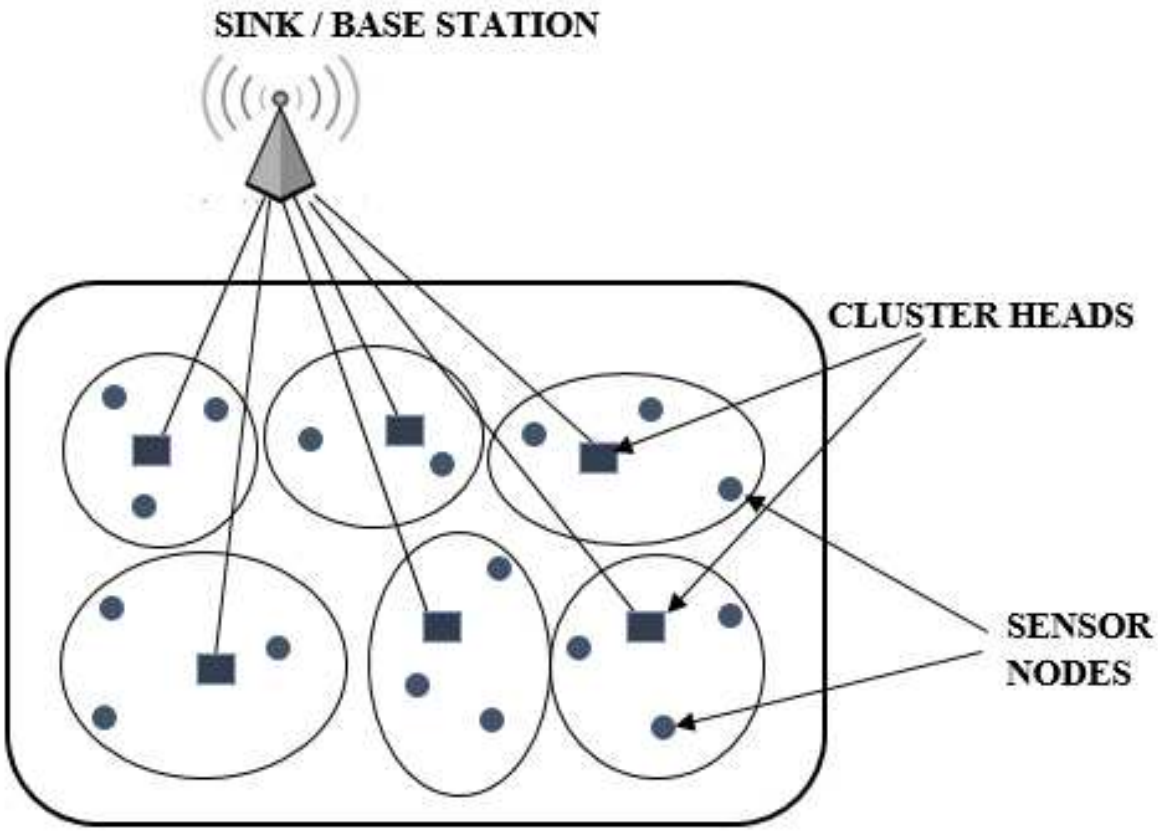

Figure: A cluster-based wireless sensor network

Statistical Model

- We consider N sensor nodes located at different parts of a patient's body, each measuring one of the K related biomarkers (features) at T different discrete time points.
- Each sensor node measures one particular biomarker (of interest) at T discrete time points. The biomarkers are correlated in the sense that a high (or low) value of one biomarker at time t can affect the values of the other biomarkers (some or all) at time $t+1$.
- The sensor nodes report only the binary/ordinal outcomes, say, 1 (high) and 0 (fair) based on some prefixed thresholds for each biomarker.
- The non-medical persons can better understand the patient's condition in this way. Additionally, this approach is energy efficient.

Statistical Model

- Let $Y_{ik}(t)$ be the binary outcome at time t obtained from the i-th sensor node measuring the k-th biomarker (and hence belonging to the k-th cluster).

- At the base station, we only receive $Y_{ik}(t)$, and the corresponding unobserved measurement $Y^*_{ik}(t)$ is considered as a latent random variable.

- The latent or unobserved continuous variables $Y^*_{ik}(t)$ and the observed $Y_{ik}(t)$ are related as follows:

$$Y_{ik}(t) = \begin{cases} 1, & \text{for } Y^*_{ik}(t) > c_k; \\ 0, & \text{for } Y^*_{ik}(t) \leq c_k, \end{cases}$$

for the prefixed (known) constants c_k, $k = 1, 2, \ldots, K$.

- The above approach is similar to Bayesian data-augmentation method proposed in Albert and Chib [1993].

Statistical Model

- For the latent random variables, we consider:

$$Y^*_{ik}(t) = f_k(t) + \alpha_k Y^*_{ik}(t-1) + \beta_k Z^*_{ik}(t-1) + \gamma_{ik} + e_{ik}(t), \quad (1)$$

where, $Z^*_{ik}(t-1)$ denotes the average Y^* values at time $t-1$ from all the other $n-1$ nodes (except the i-th node) measuring the k-th biomarker.

- Here, γ_{ik}s denote the sensor and biomarker specific random effects; and the residual errors $e_{ik}(t)$s are assumed to be identically and independently distributed as the Gaussian distribution with mean=0 and variance=σ^2.

- The function f_k essentially captures the general effect of time on the k-th biomarker, and is modeled as:
$f_k(t) = \delta_{0k} + \delta_{1k}t + \delta_{2k}t^2 + \ldots + \delta_{r_k k}t^{r_k}$.

Statistical Model

- In (1), the random effects γ_{ik} are used for capturing the correlations (dependence) among the biomarkers. We assume $\boldsymbol{\gamma}_i = [\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{iK}]^T$ are identically and independently distributed as a K-variate Gaussian distribution with mean=0 and unknown covariance matrix=Σ.

- The correlation between $Y^*_{ik}(t)$ and $Y^*_{ik'}(t)$ given all the measurements till time $t-1$ can be expressed as: $Cor(Y^*_{ik}(t), Y^*_{ik'}(t)) = \frac{Cov(\gamma_{ik}, \gamma_{ik'})}{\sqrt{Var(\gamma_{ik}) Var(\gamma_{ik'})}}$, where $Cov(\gamma_{ik}, \gamma_{ik'})$ is the (k, k')-th element of Σ; and $Var(\gamma_{ik})$ and $Var(\gamma_{ik'})$ are the k-th and k'-th diagonal element of Σ respectively.

- Thus, the above model captures the inter-biomarker dependence and intra-biomarker dependence simultaneously.

Bayesian Computations

- We consider a Bayesian approach and estimate the coefficients by Markov Chain Monte Carlo (MCMC) method.
- We consider a multivariate ($r_k + 1$ variate) Gaussian prior with mean=0 and covariance matrix=$\sigma^2_{\delta_k} I$, for $\boldsymbol{\delta}_k = [\delta_{0k}, \delta_{1k}, \ldots, \delta_{r_k k}]^T$. An Inverse Gamma (κ_1, κ_2) prior is taken for σ^2 and a Wishart (V, p) prior is taken for the matrix Σ^{-1}.
- For $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we consider diffuse priors simply because it is quite unlikely to have some prior information on these coefficients.
- Define the set of unknown coefficients $\boldsymbol{\theta} = [\boldsymbol{\delta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \Sigma]$. We need to estimate $\boldsymbol{\theta}$ based on the available data.

Posterior Distribution

- The likelihood of $\mathbf{Y}^*$ conditional on $\mathbf{Y}$ and $\boldsymbol{\theta}$ can be expressed as the following:

$$L(\mathbf{Y}^*|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\gamma}) = \prod_{k=1}^{K}\prod_{i=1}^{n}\prod_{t=1}^{T}\{1(Y_{ik}^*(t) > c_k)1(Y_{ik}(t)=1) \\ +1(Y_{ik}^*(t) \le c_k)1(Y_{ik}(t)=0)\} \times \phi(Y_{ik}^*(t)|\gamma_i), \tag{2}$$

  where, $\phi(Y_{ik}^*(t)|\gamma_i)$ denotes the Gaussian density of $Y_{ik}^*(t)$ conditional on γ_i from (1).

- Using Bayes theorem, joint posterior distribution is:

$$\pi(\boldsymbol{\theta},\boldsymbol{\gamma},\mathbf{Y}^*|\mathbf{Y}) \propto L(\mathbf{Y}^*|\mathbf{Y},\boldsymbol{\theta},\boldsymbol{\gamma}) \times \prod_i g(\boldsymbol{\gamma}_i) \\ \times\pi(\boldsymbol{\delta}) \times \pi(\boldsymbol{\alpha}) \times \pi(\boldsymbol{\beta}) \times \pi(\sigma^2) \times \pi(\Sigma), \tag{3}$$

  where g denotes multivariate normal density of $\boldsymbol{\gamma}_i$.

Estimation and Prediction

- The Gibbs sampler (based on MCMC iterations) is a very efficient computational tool for estimating the model parameters.
- We skip the computational details, and just note that it takes 2-3 minutes to estimate all the model parameters using Gibbs sampler.
- Based on the estimated model (1), we predict the latent variables, and hence get the binary predictions on the patient's health condition for the future time points.
- The predicted outcomes are sent (through internet) to the healthcare providers, and also to the patient's family.

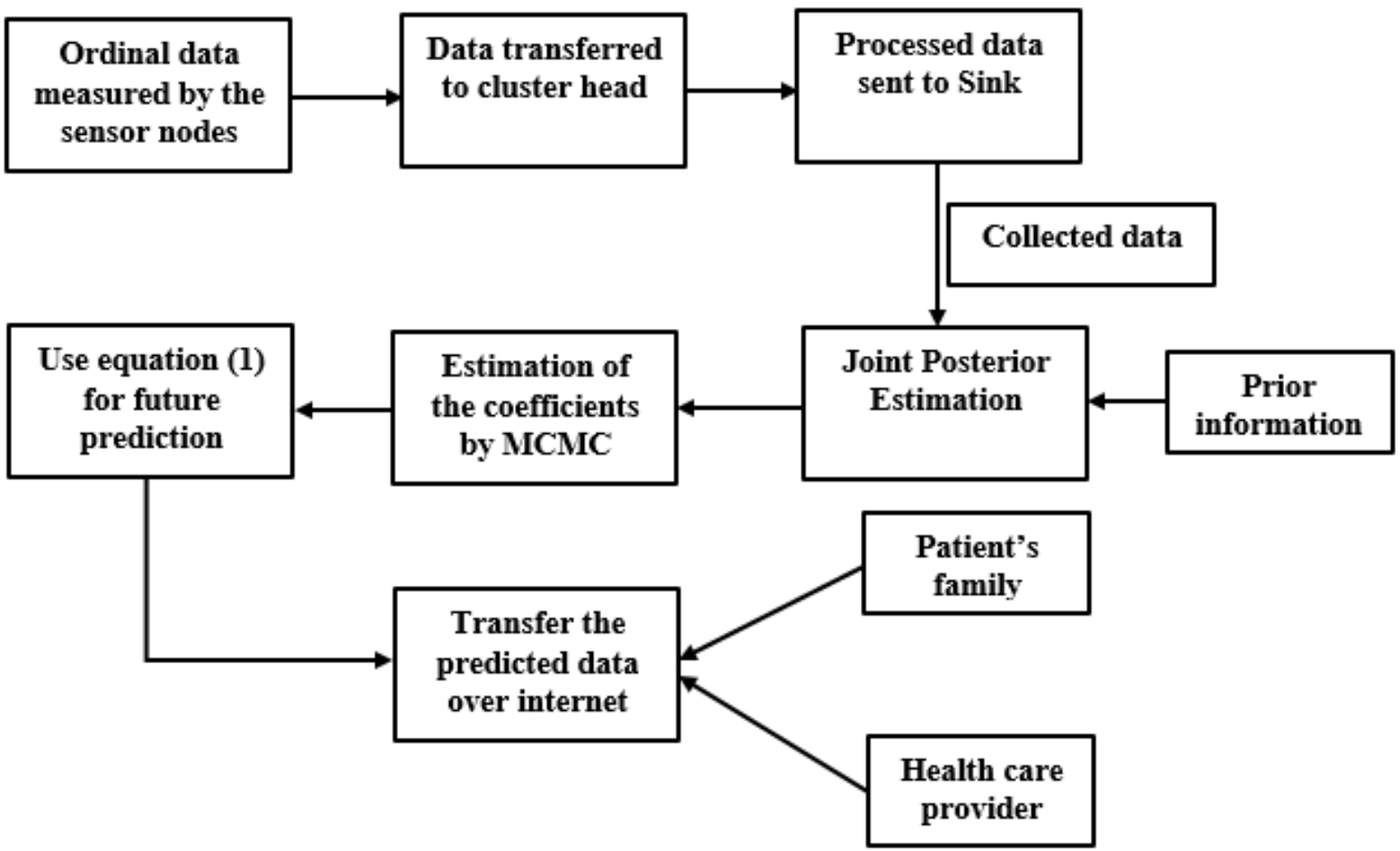

Figure: A flowchart summarizing the proposed patient monitoring approach.

Simulation Study: 1

- We consider N=30 sensor nodes (10 sensors for each of the 3 biomarkers) measuring 3 biomarkers longitudinally at T=12 evenly spaced time points.

- We generate latent random variables from the following linear model:

$$Y^*_{ik}(t) = \delta_{0k} + \delta_{1k}t + \alpha_k Y^*_{ik}(t-1) + \beta_k Z^*_{ik}(t-1) + \gamma_{ik} + e_{ik}(t), \quad (4)$$

where, δ_{01}=1.15, δ_{11}=-0.78, δ_{02}=2.79, δ_{12}=-1.65, δ_{03}=1.76, δ_{13}=-0.87. We take $\alpha_1 = 1.5$, $\alpha_2 = 2$, $\alpha_3 = 2.8$; and $\beta_1 = 0.86$, $\beta_2 = 1.6$, $\beta_3 = 1.3$. Assume $e_{ik}(t) \sim N(0,1)$ distribution, are iid.

- We generate $\gamma_i = [\gamma_{i1}, \gamma_{i2}, \gamma_{i3}]^T$ from $N_3(0, \Sigma)$, where $\Sigma = \begin{bmatrix} 4 & 3.6 & 5.6 \\ & 9 & 9.6 \\ & & 16 \end{bmatrix}$. This assigns the value of the correlation coefficients between the responses (1,2), (1,3) and (2,3) as 0.6, 0.7 and 0.8 respectively.

Simulation Study: 1

- Once the latent variables are simulated, we generate the observed binary responses $Y_{ik}(t)$ based on some prefixed thresholds ($c_1 = 10, c_2 = 15, c_3 = 8$). Now we treat the binary outcomes as the observed responses and fit the model given in equation (1).
- We use first 10 time points for model fitting and the last 2 time points for the prediction purpose.
- We fit the model given in (1) and then predict latent responses at t=11 and 12. Based on the prefixed thresholds, we obtain the predicted binary responses for t=11, and 12. We repeat this for $N = 60, 150, 300$; and compute misclassification proportion based on 100 replications (1 misclassified as 0, or vice versa).

Results

Table: Misclassification proportions for different responses for different sample sizes in Simulation I.

	Biomarker 1		Biomarker 2		Biomarker 3	
N	t=11	t=12	t=11	t=12	t=11	t=12
9	0.08	0.12	0.07	0.11	0.10	0.13
15	0.06	0.09	0.06	0.10	0.09	0.11
24	0.05	0.08	0.06	0.07	0.05	0.08
30	0.05	0.07	0.06	0.06	0.04	0.08

Simulation Study: 2

- Sometimes it is not meaningful to consider only two ordinal categories for the biomarkers. Depending on the patient's health condition, we may want to monitor the patient more critically.

- We perform simulation study to assess the performance of our predictive model when some of the biomarkers of interest have more than 2 ordinal categories. We consider 3 biomarkers, each of which is measured by 5 sensor nodes over 10 time points.

- First we simulate $Y^*_{ik}(t)$ as in study 1, and obtain the multinomial responses $Y_{ik}(t)$ as the following:
  For k=1,2;

$$Y_{ik}(t) = \begin{cases} 2, & \text{for } Y^*_{ik}(t) > c_{k_2}; \\ 1, & \text{for } c_{k_2} > Y^*_{ik}(t) > c_{k_1}; \\ 0, & \text{for } Y^*_{ik}(t) \leq c_{k_1}. \end{cases}$$

Simulation Study: 2

- For k=3;
  $$Y_{ik}(t) = \begin{cases} 1, & \text{for } Y^*_{ik}(t) > c_k; \\ 0, & \text{for } Y^*_{ik}(t) \leq c_k; \end{cases}$$
  where, c_{k_l} for l=1,2; k=1,2; and c_k for k=3; are known constants.

- We fit our model on the generated multinomial responses for the first 5 time points to estimate the regression coefficients. We use these coefficients to predict the condition of the patient (with respect to different biomarkers) at time point 6 and find out misclassification proportion.

- We keep on repeating this by including data on one more time point in a roll-in manner, and classify the patient's condition for the next time point, and compute the misclassification proportion.

Results

Table: Estimated misclassification error rates for different biomarkers over time in Simulation 2.

Time point	B-1	B-2	B-3
6	0.15	0.0	0.13
7	0.12	0.08	0.10
8	0.10	0.05	0.05
9	0.04	0.03	0.02
10	0.03	0.01	0.01

Real data analysis

- We focus on monitoring and predicting three important biomarkers related to cardiovascular disorder of a patient, namely, systolic blood pressure (normal range 120-140), diastolic blood pressure (normal range 70-90), and heart rate (normal range 72-85).

- We consider 10 sensor nodes for the heart rate, 5 sensor nodes for the diastolic BP; and 5 sensor nodes for the systolic BP. Three biomarkers are measured at 10 discrete time points, not necessarily evenly spaced.

- For systolic BP, a response $>$150 is treated as high (coded by 2); $<$100 is recorded as low (coded by 0); and a value in-between is recorded as fair (coded by 1). For the diastolic BP, similar thresholds are 90, and 70, respectively. For the heart rate we take 85 as our threshold; i.e. a value > 85 is recorded as high (coded by 1), and < 85 is recorded as normal (coded by 0).

Real data analysis

- First, we consider the data for the first 6 time points, and use this as the training data. The remaining part of the data is used as the test data.
- We use the estimated regression coefficients in equation (1), and predict the latent variable Y^* for each sensor node at T=7. The predicted latent variables are then categorized using the threshold values mentioned earlier. We compute the misclassification proportions for each biomarker. We repeat this for $T = 8, 9, 10$ and the results are summarized.
- We notice that the misclassification proportions are mostly 0 for T=7,8; and for all the 4 time points, the maximum misclassification proportion is 0.2. This illustrates that our approach provides reasonably good predictions.

Results

Table: Estimated misclassification proportions for the proposed joint model.

Time point	Systolic BP	Diastolic BP	Heart Rate
7	0	0	0.1
8	0	0	0
9	0.2	0.2	0
10	0	0.2	0.1

Comparison with Other Models

- We also analyze the same data using two existing approaches, namely, (i) state-space model for dynamic state estimation, and (ii) separate linear mixed model for each biomarker.

- Traditionally, state-space models are used for dynamic state estimation in wireless communications. We consider the following model for our latent variables:

$$Y_{ik}^*(t) = Z_{ik}^*(t) + e_{ik}(t), \quad Z_{ik}^*(t) = \alpha Z_{ik}^*(t-1) + \epsilon_{ik}(t), \tag{5}$$

  where, $e_{ik}(t)$ and $\epsilon_{ik}(t)$ independently follow $N(0, \sigma_e^2)$ and $N(0, \sigma_\epsilon^2)$ respectively.

- We also need to consider $|\alpha| < 1$ for valid inference. We estimate the regression coefficients using Kalman-Filter.

Comparison with Other Models

- Next, we consider the linear mixed models for each biomarker separately. In other words, we do not consider correlations among the biomarkers and estimate the latent responses independently.
- Thus, we use the following model:

$$Y^*_{ik}(t) = f_k(t) + \alpha_k Y^*_{ik}(t-1) + \beta_k Z^*_{ik}(t-1) + e_{ik}(t). \tag{6}$$

- This model is similar to Chatterjee et al. [2016] where they use Gibbs sampler for state estimation and anomaly detection. Note that such separate modeling assumes that the biomarkers are independent, and hence one biomarker does not affect the others.

Results

Table: Estimated misclassification proportions for dynamic state-space model, and separate linear mixed models.

	dynamic state-space model			separate linear models		
Time point	SBP	DBP	HR	SBP	DBP	HR
7	0.2	0.2	0.1	0.2	0.2	0.1
8	0.2	0.4	0.1	0.4	0.4	0.2
9	0.4	0.4	0.1	0.2	0.6	0.2
10	0.4	0.6	0.2	0.4	0.4	0.3

Summary

- We propose a flexible Bayesian dynamic joint model which can effectively monitor multiple biomarkers over time, and can also predict the biomarkers.
- Our model is useful for monitoring patients with chronic diseases in the rural as well as in the urban areas.
- There might be some latent clusters among the patients, and based on the observed data we can use a functional clustering method for detecting such latent clusters.
- Cluster-based modeling should improve the predictive power of the model.

Acknowledgement

- This is a joint work with two of my collaborators.
- My former Ph.D student, Dr. Jayabrata Biswas, Department of Statistics, University of Burdwan, India.
- Dr. Aditi Chatterjee, Department of Electronics & Telecommunications Engineering, Jadavpur University, India.

References

- Albert J, Chib S. Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association 1993; 88: 669-679.
- Chatterjee A, Venkateswaran P, Das K. Simultaneous State Estimation for Clustered Based Wireless Sensor Networks. IEEE Transactions on Wireless Communications 2016; 15: 7985–7995.
- Chatterjee A, Biswas J, and Das K. An automated patient monitoring using discrete-time wireless sensor networks. International Journal of Communication Systems 2020, https://doi.org/10.1002/dac.4390.

Thank You