Switching-state Dynamical Modeling of Daily Behavioral Data

Randy Ardywibowo

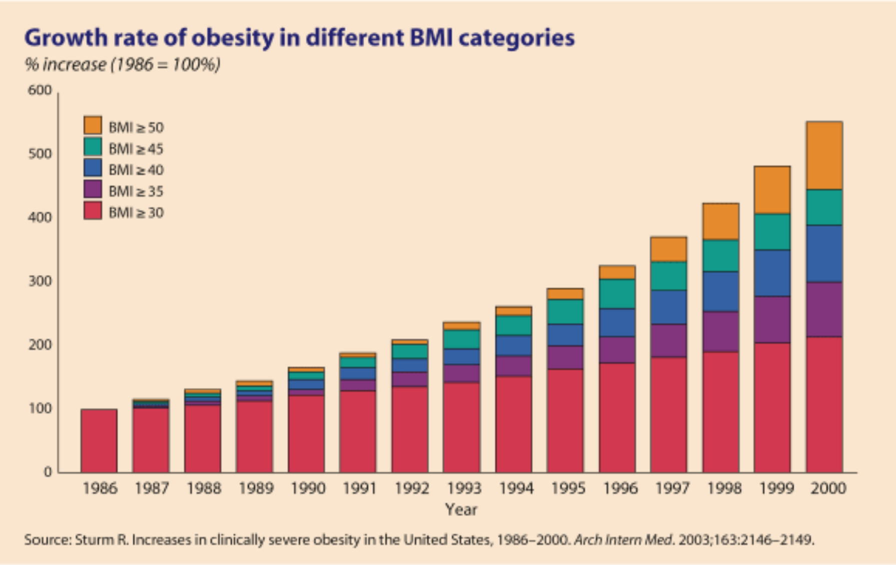
Shuai Huang, Cao Xiao, Shupeng Gui, Yu Cheng, Ji Liu, Xiaoning Qian

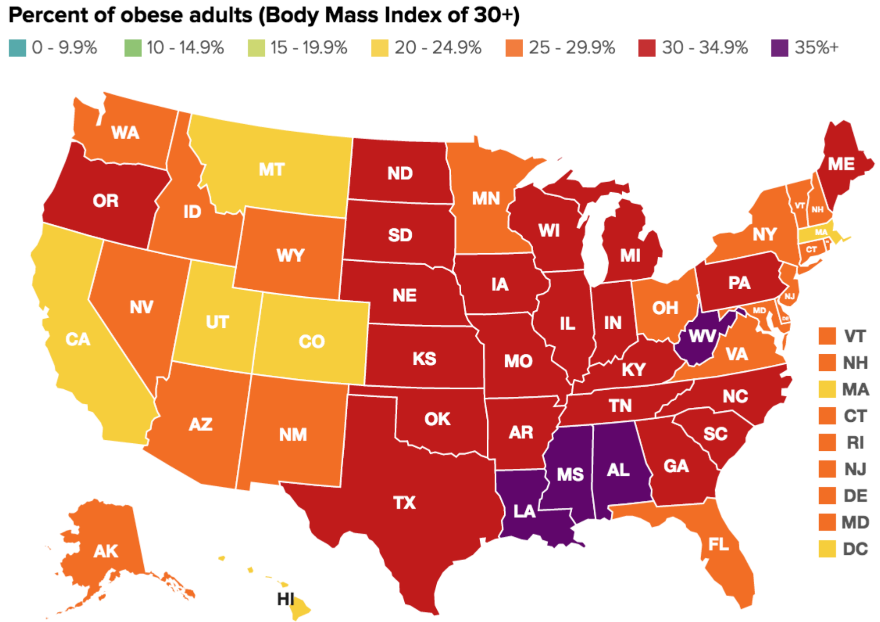
Texas A&M University, University of Washington, IBM T.J. Watson Research Center, University of Rochester

6th Workshop on Data Mining for Medicine & Healthcare

Background

- Obesity is a public health issue. Currently, over 1/3 of U.S. population is obese.
- Beyond the capacity of the healthcare industry.
- For obesity, treating patients in hospitals full time is not a possibility.
- Motivates development of scalable solutions to control obesity.





Adult Obesity in the U.S. September 2016
stateofobesity.org

Background

- Smartphones & sensors provide unprecedented health monitoring capabilities.
- Enables more personalized and fine-grained health intervention and treatments.
- Can we develop smart, scalable solutions that can automate personalized monitoring, intervention, and activity planning?

Example of a Health Monitoring App

apple.com/ios/health/

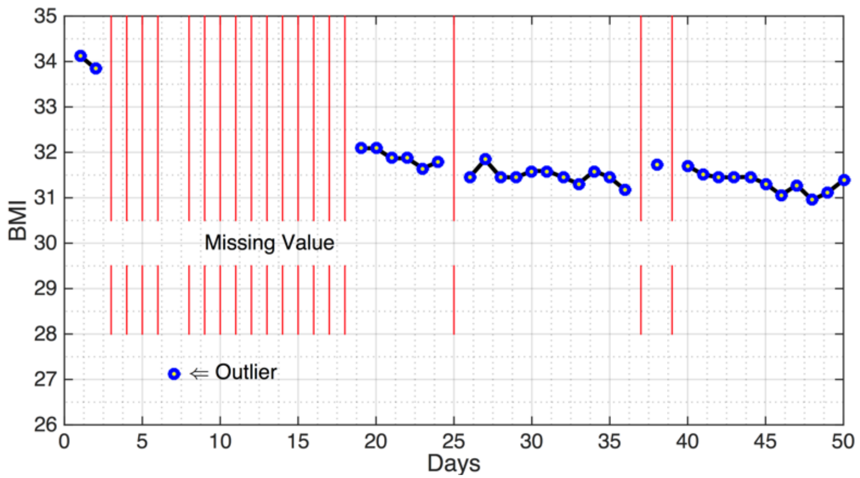
Example of wearable sensors

theverge.com

Challenges

- Complex daily life behaviors.
  - Different people at different times react to food intake, exercise, and other behaviors differently.
  - Complex relationships between behavior variables and outcome variables.
  - Model heterogeneity while accounting for similarities between different people.

- Missing values and outliers in data.
  - Defects abundant in healthcare data.
  - Noisy and irregular wearable sensor data.



Daily life behavior data from sensors

Related Work

- Dynamical systems modeling
  - Applications: robot control, weather prediction, market prediction, health monitoring.
  - Techniques: State-space methods, Spectral analysis, Auto-Regressive Moving Average (ARMA).

- Missing value and outlier treatment
  - Off-the-shelf imputation methods.
  - Functional data analysis.

Outline

- ~~Background~~
- ~~Related Work~~
- **Population Switching-state Auto-Regressive (SAR) Model**
- SAR Model Learning
- Simultaneous Missing Value and Outlier Treatment
- Evaluation
- Model Interpretation
- Summary

Population SAR Model

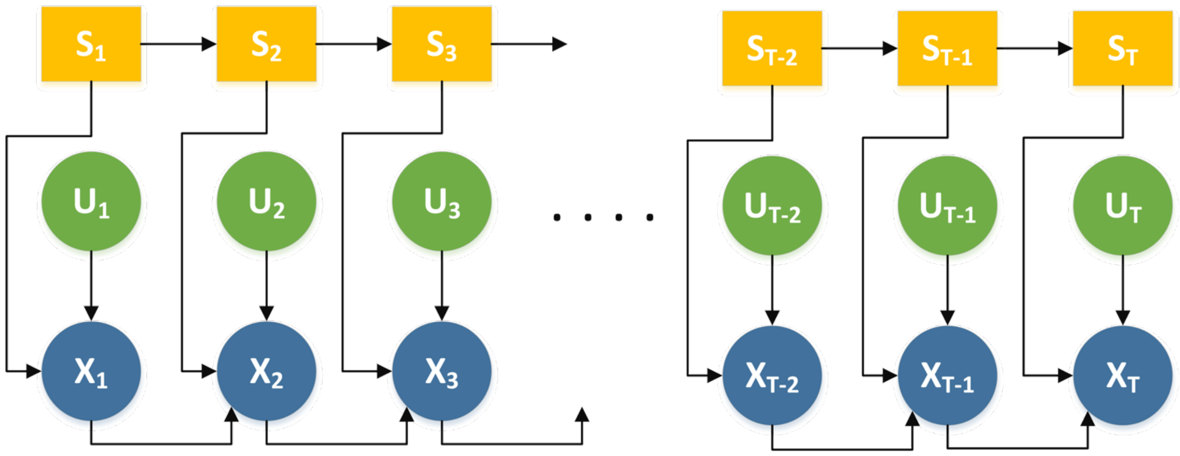
- Hidden Markov Model (HMM).
- s_t: Hidden States. Models the dynamic **heterogeneity**.
- x_t: Observed health indicator, Body Mass Index (BMI).
- u_t: Input actions (Workout, exercise, food consumption).
- t: Time Index; i: Subject index.

- *Parameters*: L_x for $\mathbf{x}_{t-1}^i$, L_u for $\mathbf{u}_t^i$, and S for number of states.

$$x_t^i = \boldsymbol{a}(s_t^i)^{\mathrm{T}}\mathbf{x}_{t-1}^i + \boldsymbol{b}(s_t^i)^{\mathrm{T}}\mathbf{u}_t^i + c(s_t^i) + \varepsilon_i$$

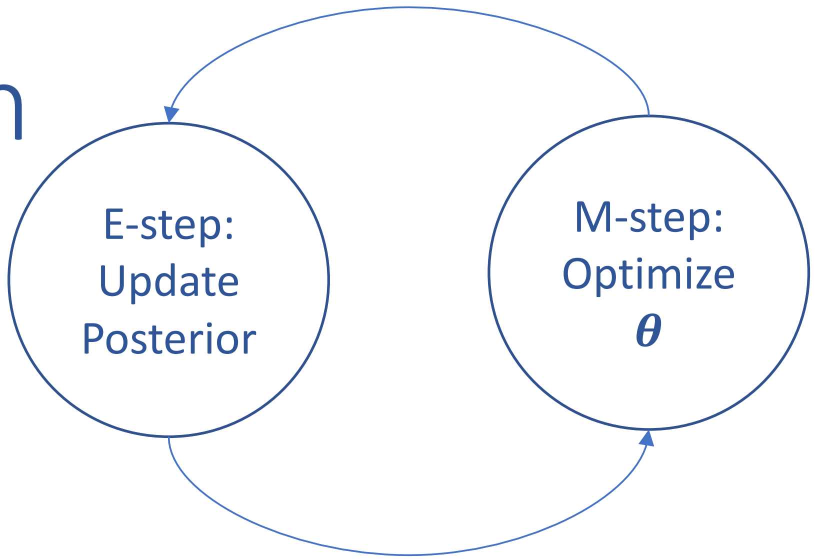
$$\varepsilon_i \sim N(0, \sigma_i(s))$$

$\boldsymbol{a}(s)$, $\mathbf{b}(s)$, $\mathrm{c}(s)$ shared between subjects



SAR Model for L_x = L_u = 1

Maximum Likelihood Estimation



- Learn $\boldsymbol{\theta} = \{\mathbf{a}(s), \mathbf{b}(s), c(s), \sigma_i^2(s), \forall s \in \{1, \dots, S\}\}$.
- Expectation-Maximization Algorithm with hidden variable s.
- Minimize the Kullback-Leibler (KL) divergence.
- Denote:

$$\mathbf{d}(s_t^i) = \begin{bmatrix} \mathbf{a}(s_t^i) \\ \mathbf{b}(s_t^i) \\ c(s_t^i) \end{bmatrix} \quad \hat{\mathbf{v}}_{t-1}^i = \begin{bmatrix} \hat{\mathbf{x}}_{t-1}^i \\ \hat{\mathbf{u}}_t^i \\ 1 \end{bmatrix}$$

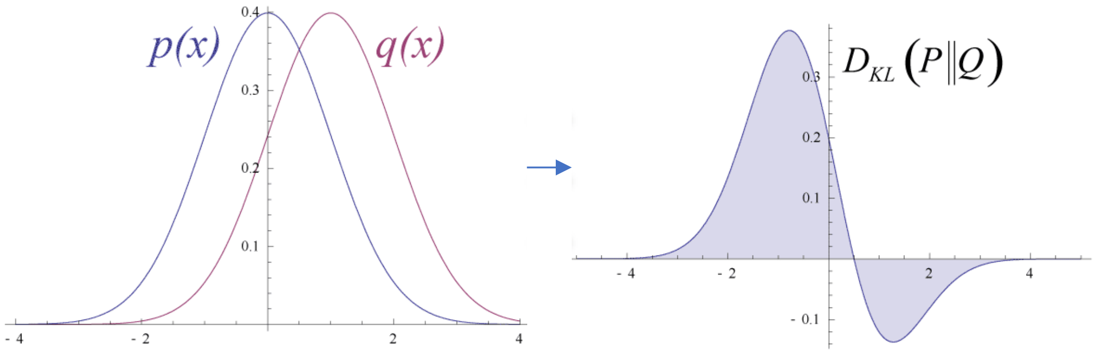


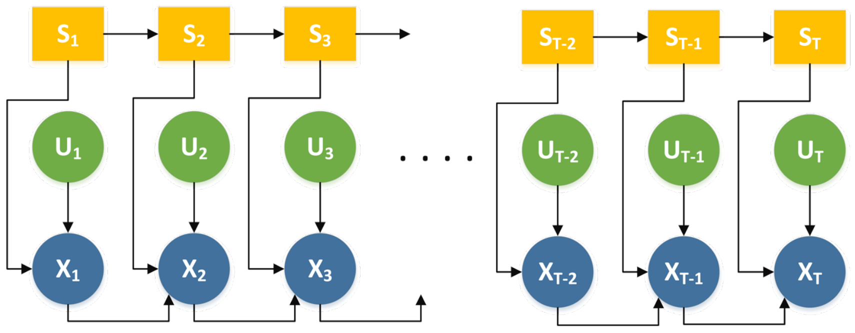
Illustration of the KL divergence

wikipedia.org

$$E = \sum_i \sum_t \langle \log p(x_t^i | \hat{\mathbf{v}}_{t-1}^i, \mathbf{d}(s_t^i)) \rangle_{p^{old}(s_t^i | x_{1:T}^i)} + \sum_i \sum_t \langle \log p(s_t^i | s_{t-1}^i) \rangle_{p^{old}(s_t^i, s_{t-1}^i)}$$

E-step

- Find posterior $\gamma(s_t^i) = p(s_t^i | x_{1:T}^i, u_{1:T}^i, \boldsymbol{\theta})$.
- Dynamic programming approach using the *Forward-Backward Algorithm*.
- Recursive message passing from both future (backwards) and past (forward) information.
- Forward pass:
  - $\alpha(s_t^i) = \underbrace{p(x_t^i | s_t^i, \mathbf{x}_{t-1}^i, \mathbf{u}_t^i, \boldsymbol{\theta})}_{\text{Corrector}} \underbrace{\sum_{s_{t-1}^i} p(s_t^i | s_{t-1}^i)\, \alpha(s_{t-1}^i)}_{\text{Predictor}}$



- Backward pass:
  - $\beta(s_{t-1}^i) = \sum_{s_t^i} p(x_t^i | s_t^i, \mathbf{x}_{t-1}^i, \mathbf{u}_t^i, \boldsymbol{\theta}) p(s_t^i | s_{t-1}^i)\, \beta(s_t^i)$
  - "Normalizes past observations based on the trajectory actually seen."

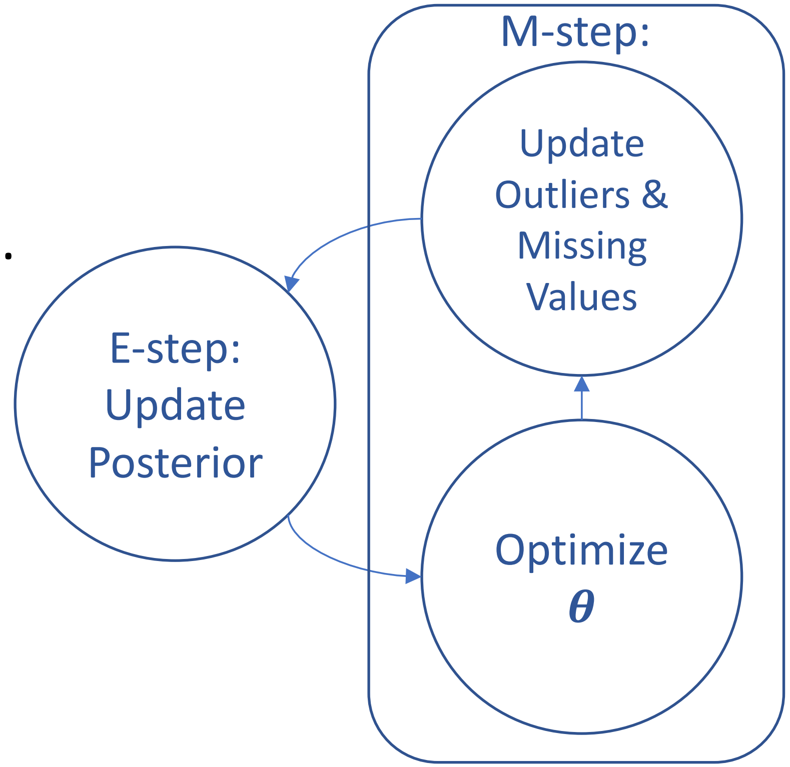
- $\gamma(s_t^i) = \dfrac{\alpha(s_t^i)\beta(s_t^i)}{\sum_{s_t^i} \alpha(s_t^i)\beta(s_t^i)}$

M-step with Simultaneous Missing Value and Outlier Treatment

- Optimize $\boldsymbol{\theta}$ to the given the posterior probabilities.
- Minimize the Kullback-Leibler (KL) divergence.

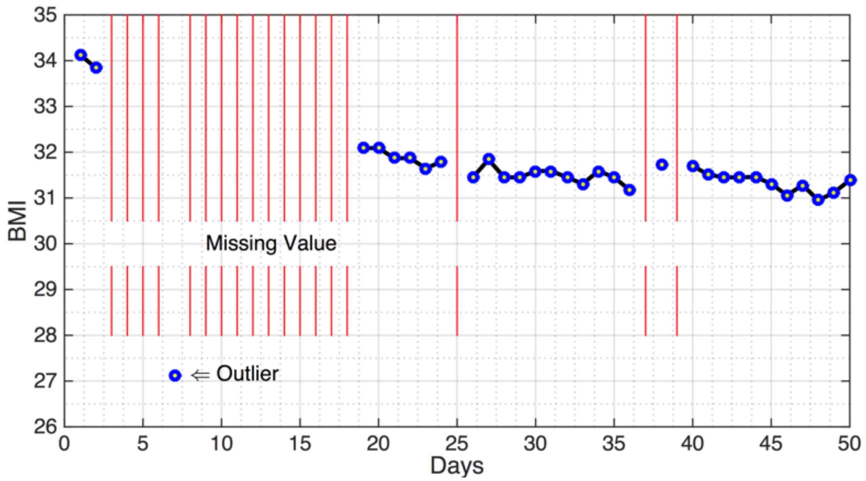
- Missing values and outliers should best fit the learned model.
  - $\min_{\hat{X},\hat{U}} \sum_{i=1}^{N} \sum_{t=0}^{T-1} \left\| x_t^i - \mathbf{d}\left(s_t^i\right)^{\mathrm{T}} \hat{\mathbf{v}}_{t-1}^i \right\|^2$

- Define outliers as the top η_x and η_u deviate points in $\widehat{X},\ \widehat{U}$.
- Limit the number of outliers at each EM episode.
  - $\text{s.t.} \left\| \left(\hat{X}_i - X_i\right)_{\Omega_{x_i}} \right\|_0 \le \eta_x, \left\| \left(\widehat{U}_i - U_i\right)_{\Omega_{u_i}} \right\|_0 \le \eta_u$



Dataset

- Daily behavioral data collected from sensors and patient logs.
- 25 different subjects.
- Consists of workout calories burned, exercise calories, workout time, and calories consumed.
- Contains severe missing value and outlier patterns.



Daily life behavior data from sensors

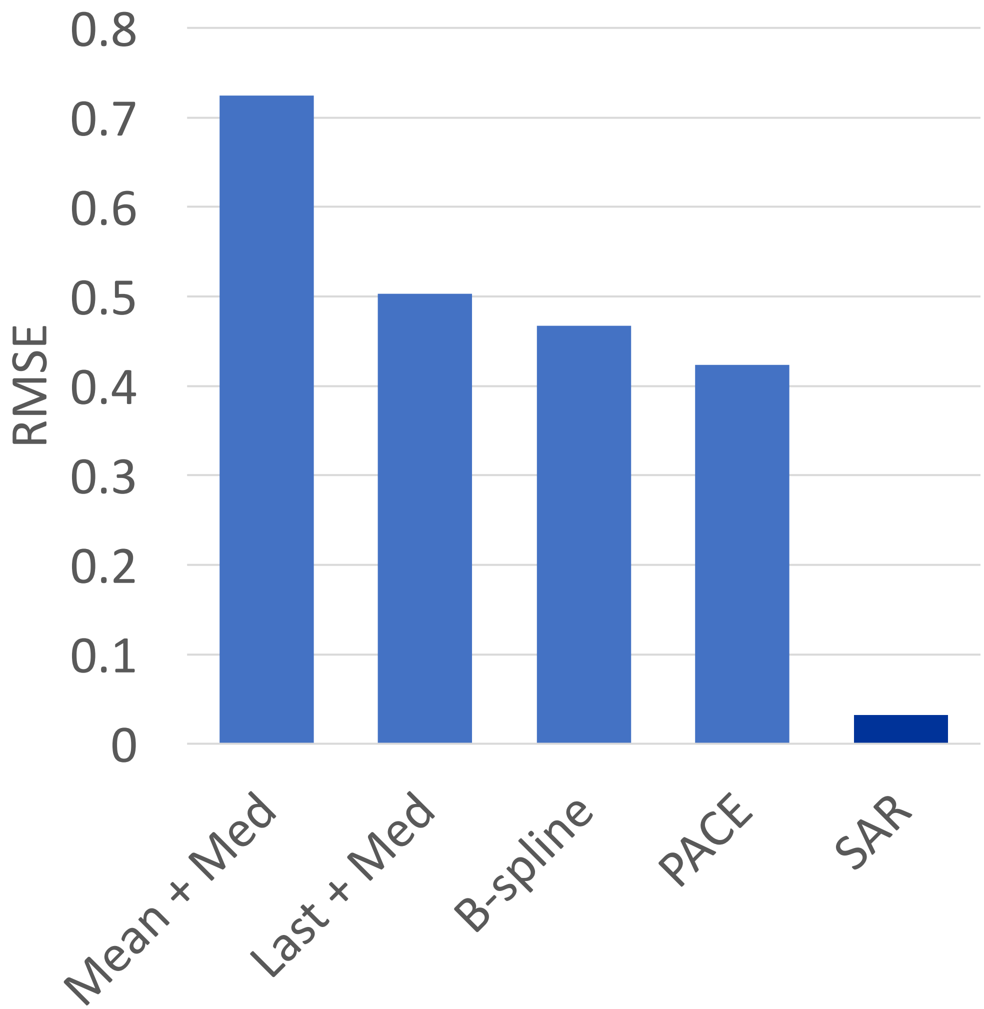
Missing Value and Outlier Treatment Evaluation

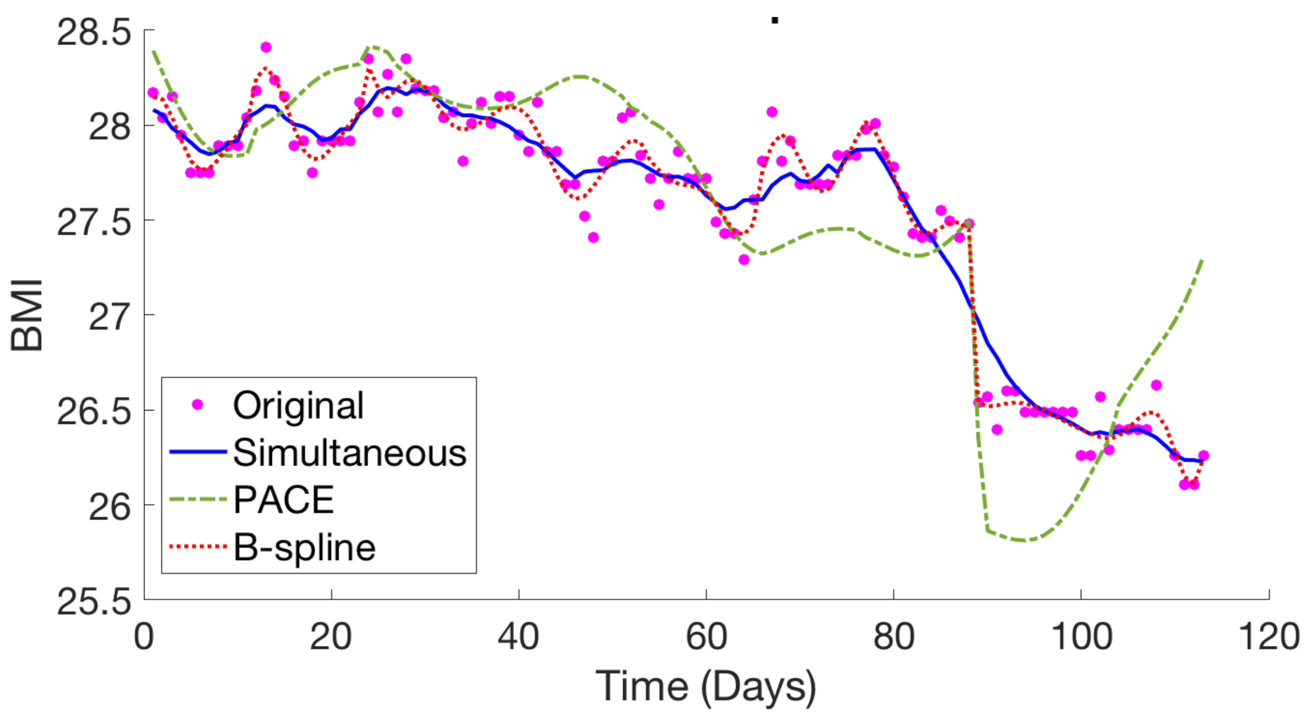
- Compare with off-the-shelf missing value and outlier detection
  - Mean, Last-value-carried-forward, Median filter.

- Compare with functional data analysis (FDA).
  - **PACE** (Functional Principal Component Analysis through Conditional Expectation) and FPCA with B-spline.

	Mean + Med	Last + Med	B-spline	PACE	Simultaneous
RMSE	0.72463 ± 0.53953	0.50286 ± 0.48841	0.46668 ± 0.48512	0.42348 ± 0.48489	**0.032114 ± 0.016565**
ABS	0.29995 ± 0.17872	0.17697 ± 0.12145	0.15886 ± 0.12179	0.11965 ± 0.10518	**0.024125 ± 0.011585**

Missing Value and Outlier Treatment Evaluation



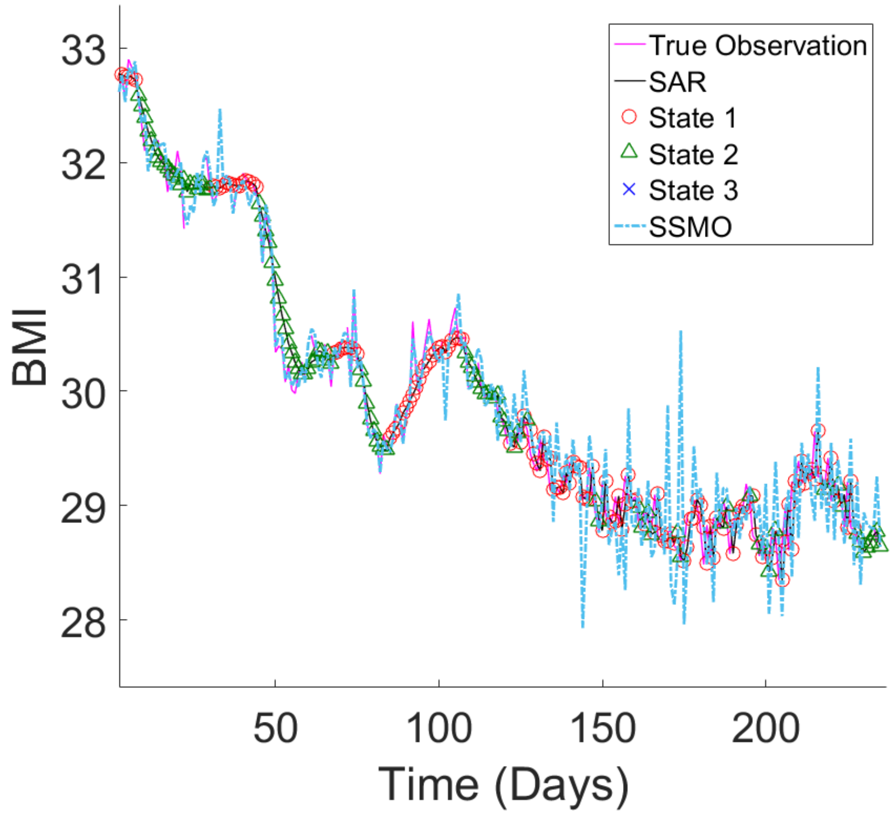


Switching and Population Effects Evaluation

- Compare with a similar model without switching and population effects, denoted as **SSMO**.

- Separate $\boldsymbol{\theta} = \{A, B, C\}$ for each subject.

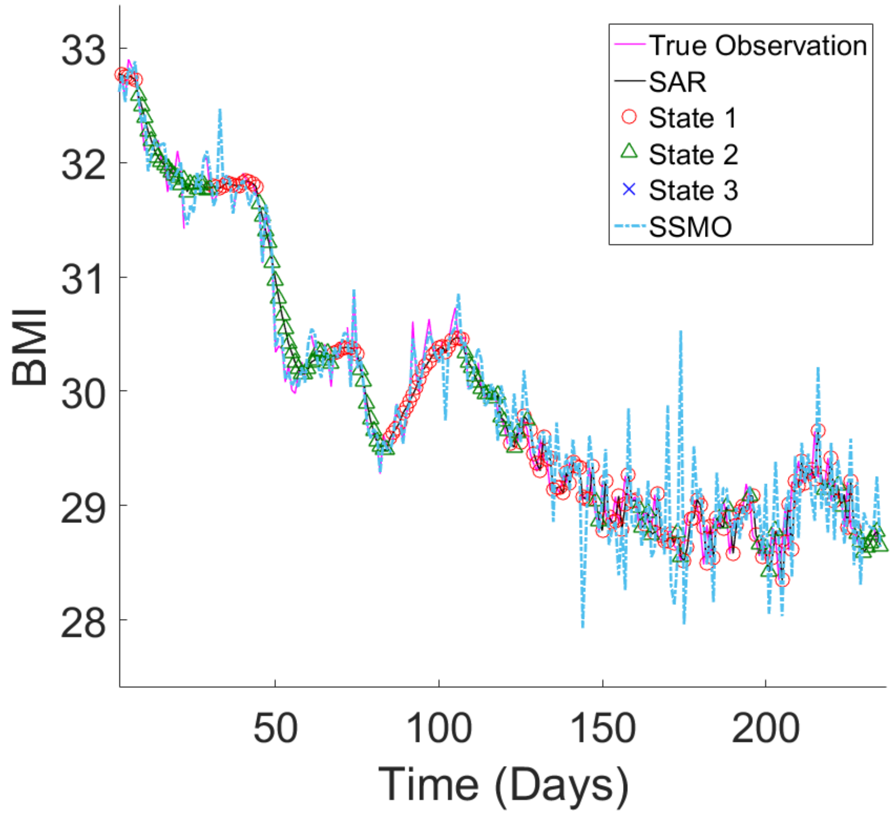
	SSMO	SAR
ABS	0.22353 ± 0.28732	**0.024125 ± 0.011585**
RMSE	0.40324 ± 0.61404	**0.032114 ± 0.016565**



Model Interpretation

- Best parsimonious model: $S = 3, L_x = 1, L_u = 1.$
- State 1: Least active.
- State 2: Most active.
- State 3: Intermediate state.

Variable	State 1	State 2	State 3
BMI	1.0003	0.9824	0.9950
Exercise Calories	-0.0032	-0.0047	-0.0043
Food Calories	0.0007	0.0187	0.0104
Workout Calories	0.0031	-0.0080	0.0017
Workout Time	-0.0251	0.0261	0.0072



Summary

- A Population SAR model was implemented on daily behavioral data.
- The EM learning procedure simultaneously treats missing values and outliers.
- The best parsimonious model selected was where $S = 3$, $L_x = L_u = 1$.
- The method outperforms missing value and outlier preprocessing methods.
- Taking into account the dynamic heterogeneity and population effects improves model accuracy significantly.

Future Work

- Intervention and control.
  - Standard techniques: Reinforcement learning, Markov Decision Process (MDP).
  - Underdetermined problem. Many options for intervention.
  - Can we learn the model and the control policy **simultaneously**?

- Intervention evaluation.
  - How can we evaluate the quality of intervention?
  - Clinical trials are costly and needs approval.
  - **Model checking**: theoretical guarantees through simulation.
  - Simulation environments may not accurately model real-world usage.

Works Cited

- D. Barber, Bayesian reasoning and machine learning: Cambridge University Press, 2012.
- S. Boyd and L. Vandenberghe, Convex optimization: Cambridge university press, 2004.
- C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal, "Prevalence of childhood and adult obesity in the United States, 2011-2012," Jama, vol. 311, pp. 806-814, 2014.
- L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proceedings of the IEEE, vol. 77, pp. 257-286, 1989.
- J. O. Ramsay, Functional data analysis: Wiley Online Library, 2006.
- F. Yao, H.-G. Müller, and J.-L. Wang, "Functional data analysis for sparse longitudinal data," Journal of the American Statistical Association, vol. 100, pp. 577-590, 2005.
- C. Xiao, S. Gui, Y. Cheng, X. Qian, J. Liu, and S. Huang, "Learning Longitudinal Planning for Personalized Health Management from Daily Behavioral Data", in submission, 2016.

Thank You