CPM: A General Feature Dependency Pattern Mining Framework for Contrast Multivariate Time Series

Qingzhe Li[a], Liang Zhao[b,*], Yi-Ching Lee[a], Avesta Sassan[a], Jessica Lin[a]

[a] *George Mason University, 4400 University Drive, Fairfax, VA, USA*
[b] *Emory University, 201 Dowman Dr, Atlanta, GA 30322*

1. Model Optimization

In this section, we elaborate our optimization algorithm for the proposed CPM-C model.

Optimizing Equation (7) with covariance based regularization term
To solve optimization problem (7) with $\boldsymbol{\delta}(\cdot)$ defined in Equation (3), we first rewrite the objective function in an ADMM-friendly form with some new notations. To do so, we define $\Phi_k \in \mathbb{R}^{c(\Phi_k)\times nw}$ as all subsequences assigned to the k-th latent state without contrast patterns in X and $\hat{X}$, where $c(\Phi_k)$ is the number of subsequences in Φ_k. Similarly, we define $\Psi_k \in \mathbb{R}^{c(\Psi_k)\times nw}$ as the subsequences in $\hat{X}$ assigned to the k-th latent state with contrast patterns. We also denote $S_k = \sum_{i=1}^{c(\Phi_k)} \Phi_{k,i}^\intercal \Phi_{k,i}/c(\Phi_k)$ and $S'_k = \sum_{i=1}^{c(\Psi_k)} \Psi_{k,i}^\intercal \Psi_{k,i}/c(\Psi_k)$ as the empirical covariance matrices, where $\{S_k, S'_k\} \in \mathbb{R}^{nw\times nw}$. Let $\mathbf{Tr}(\cdot)$ be the trace of the matrix. Then the optimization problem (7) becomes:

$$\begin{aligned}
\underset{\{\theta_k,\hat{\theta}_k\}\succ 0, V_k}{\arg\min}\ & \lambda\|V_k\|_F^2 + \frac{c(\Phi_k)}{2}[\mathbf{Tr}(S_k\theta_k) - \log\det\theta_k] \\
& + \frac{c(\Psi_k)}{2}[\mathbf{Tr}(S'_k\hat{\theta}_k) - \log\det\hat{\theta}_k] \\
\text{s.t. } & \theta_k = P_k, \hat{\theta}_k = \hat{P}_k, P_kQ_k = I, \hat{P}_k\hat{Q}_k = I, V_k = Q_k - \hat{Q}_k,
\end{aligned}$$

where $\{\theta_k, \hat{\theta}_k, P_k, \hat{P}_k, Q_k, \hat{Q}_k, V_k\} \in \mathbb{R}^{nw\times nw}$.

*Corresponding author
Email address: `liang.zhao@emory.edu` (Liang Zhao)

The augmented Lagrange in scaled form is:

$$\begin{aligned}
L_\alpha = &\frac{c(\Phi_k)}{2}[\mathbf{Tr}(S_k\theta_k) - \log\det\theta_k] \\
&+ \frac{c(\Psi_k)}{2}[\mathbf{Tr}(S'_k\hat{\theta}_k) - \log\det\hat{\theta}_k] \\
&+ \lambda\|V_k\|_F^2 + \frac{\alpha}{2}(\|\theta_k - P_k + U_k^{(1)}\|_F^2 + \|\hat{\theta}_k - \hat{P}_k + U_k^{(2)}\|_F^2 \\
&+ \|P_kQ_k - I + U_k^{(3)}\|_F^2 + \|\hat{P}_k\hat{Q}_k - I + U_k^{(4)}\|_F^2 \\
&+ \|V_k - Q_k + \hat{Q}_k + U_k^{(5)}\|_F^2 \\
&- \|U_k^{(1)}\|_F^2 - \|U_k^{(2)}\|_F^2 - \|U_k^{(3)}\|_F^2 - \|U_k^{(4)}\|_F^2 - \|U_k^{(5)}\|_F^2)
\end{aligned}$$

Where $\alpha > 0$ is the ADMM penalty parameter which is typically initialized as 1, $P_k, \hat{P}_k, Q_k, \hat{Q}_k$ are auxiliary variables, $U_k^{(1)}, \cdots, U_k^{(5)}$ are the scaled dual variables with the same size of $nw \times nw$, and I is the identity matrix.

Update θ_k: The subproblem of θ_k-update is as follows:

$$\arg\min_{\theta_k \succ 0} c(\Phi_k)[\mathbf{Tr}(S_k\theta_k) - \log\det\theta_k]/2 + \frac{\alpha}{2}\|\theta_k - P_k + U_k^{(1)}\|_F^2$$

The solution is $\theta_k = D_k\Delta D_k^\intercal$, where Δ_k is a diagonal matrix whose i-th element $\Delta_{k,i,i} = (\Lambda_{k,i,i} + \sqrt{\Lambda_{k,i,i}^2 + 8\alpha c(\Phi_k)})/4\alpha$, and $D_k\Lambda_kD_k^\intercal$ is the eigendecomposition of $2\alpha(P_k - U_k^{(1)}) - c(\Phi_k)S_k$.

Update $\hat{\theta}_k$: The subproblem of $\hat{\theta}_k$-update can be solved similarly to θ_k-update. The solution is $\hat{\theta}_k = \hat{D}_k\hat{\Delta}_k\hat{D}_k^\intercal$, where $\hat{\Delta}_k$ is a diagonal matrix whose i-th element $\hat{\Delta}_{k,i,i} = (\hat{\Lambda}_{k,i,i} + \sqrt{\hat{\Lambda}_{k,i,i}^2 + 8\alpha c(\Psi_k)})/4\alpha$, and $\hat{D}_k\hat{\Lambda}D_k^\intercal$ is the eigendecomposition of $2\alpha(\hat{P}_k - U_k^{(2)}) - c(\Psi_k)S'_k$.

Update P_k and $\hat{P}_k$: The updates of P_k and $\hat{P}_k$ are solved by proximal operators in order to satisfy the positive definite constraint. The solutions are:

$$P_k \leftarrow \text{prox}_{\succeq}(\text{prox}_0((Q_k - U_k^{(3)}Q + \theta_k + U_k^{(1)})(Q_kQ_k + I)^{-1}))$$

$$\hat{P}_k \leftarrow \text{prox}_{\succeq}(\text{prox}_0((\hat{Q}_k - U_k^{(4)}\hat{Q}_k + \hat{\theta}_k + U_k^{(2)})(\hat{Q}_k\hat{Q}_k + I)^{-1}))$$

where $\text{prox}_0(x) := (0.5x + 0.5x^\intercal)_+$, and $\text{prox}_{\succeq}(A) := \sum_i^{nw}(\Lambda_{i,i})_+ D_iD_i^\intercal$ where $D\Lambda D^\intercal$ is the eigendecomposition of A.

Update Q_k and $\hat{Q}_k$: The solutions are:

$$Q_k \leftarrow (P_k + I)^{-1}(\hat{Q}_k + U_k^{(5)} + V_k - U_k^{(3)} + I)$$

$$\hat{Q}_k \leftarrow (\hat{P}_k + I)^{-1}(Q_k - U_k^{(5)} - V_k - U_k^{(4)} + I)$$

Update V_k: The solution is:

$$V_k \leftarrow \alpha(Q_k - \hat{Q}_k - U_k^{(5)})/(2\lambda + \alpha)$$

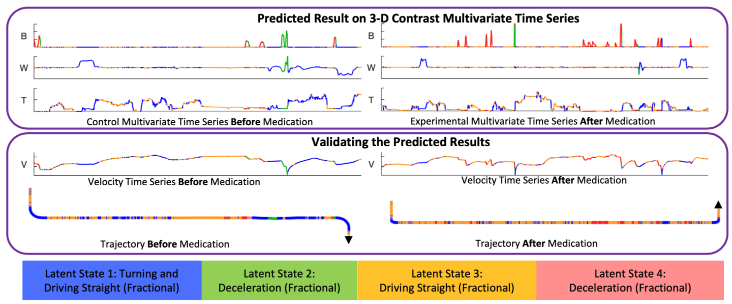
Update the dual variables $U_k^{(1)}, U_k^{(2)}, U_k^{(3)}, U_k^{(4)}, U_k^{(5)}$: The dual variables are updated as follows:

$$\begin{aligned}
U_k^{(1)} &\leftarrow U_k^{(1)} + \theta_k - P_k \\
U_k^{(2)} &\leftarrow U_k^{(2)} + \hat{\theta}_k - \hat{P}_k \\
U_k^{(3)} &\leftarrow U_k^{(3)} + P_kQ_k - I \\
U_k^{(4)} &\leftarrow U_k^{(4)} + \hat{P}_k\hat{Q}_k - I \\
U_k^{(5)} &\leftarrow U_k^{(5)} + V_k - Q_k + \hat{Q}_k
\end{aligned}$$

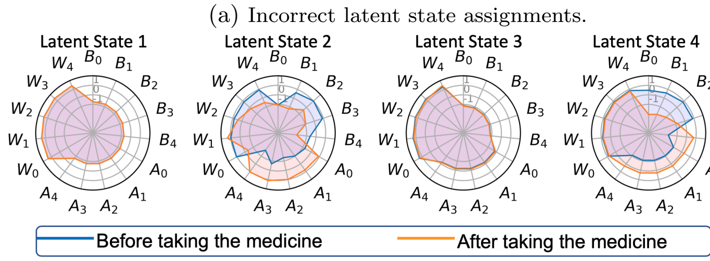
2. Generating the Synthetic Datasets

The 12 synthetic datasets are generated in 3 steps: *Step 1: Generating the ground truth latent state and CDFD pattern assignments.* Given the latent state count K, we first generate $2K$ segments by repeating each element in the list of $[1, 1, 2, 2, \cdots, K, K]$ 500 times. Then the ground truth latent state assignments are generated by randomly concatenating these $2K$ variables. We repeat the above process twice to generate the ground truth latent state assignments for the control time series (i.e.,Y) and the experimental time series (i.e.,$\hat{Y}$). To generate the ground truth CDFD pattern assignments, we randomly selected 250 consecutive points in each segment as the position with the CDFD pattern (i.e.,Z). By following this rationale, all the ground truth assignments for the 12 synthetic datasets used in this section are generated with different K values ranging from 2 to 5. *Step 2: Generating the inverse covariance matrices.* Given the dimensionality n of the multivariate time series and the sliding window size w, we generat K inverse covariance matrices $\{\theta_k \in \mathbb{R}^{nw \times nw}\}_k^K$, from K undirected Erdős-Rényi random graphs [1] with random weights. In order to maintain the fairness of the performance comparison, we follow the process described in TICC [2], which enforces the positive definite and block Toeplitz constraints on the inverse covariance matrices. To generate $\{\hat{\theta}_k\}_k^K$, we multiply a random value for each element in θ_k and then adjust the values to maintain the positive definite and block Toeplitz constraints. *Step 3: Generating the CMTS data.* We generate observations in time series from the beginning to the end. Each observation assigned to the k-th latent state in control time series is sampled from the conditional multivariate Gaussian distribution [3] parameterized by θ_k, where the condition is the previous $w - 1$ observations. Similarly, each observation assigned to the k-th latent state in the experimental time series is sampled from the conditional Gaussian distribution, which is parameterized by θ_k if this observation is labeled *without* a CDFD pattern in the ground truth assignments, or parameterized by $\hat{\theta}_k$ if this observation is labeled *with* a CDFD pattern.

3. Visualizing Baseline Results on Real-world Datasets

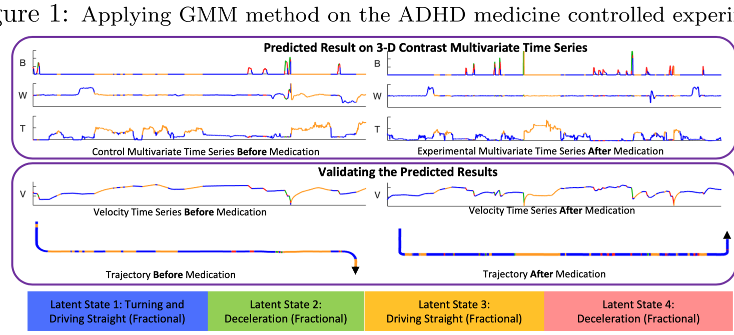


(a) Incorrect latent state assignments.

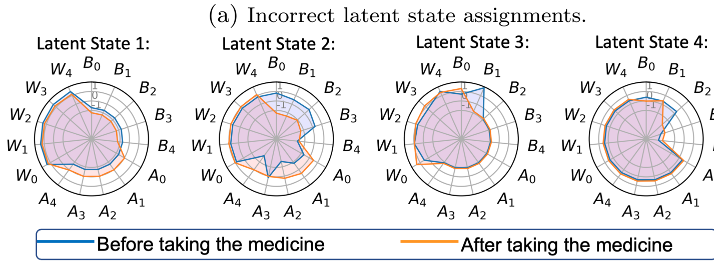


(b) Unreliable contrast patterns.

Figure 1: Applying GMM method on the ADHD medicine controlled experiment

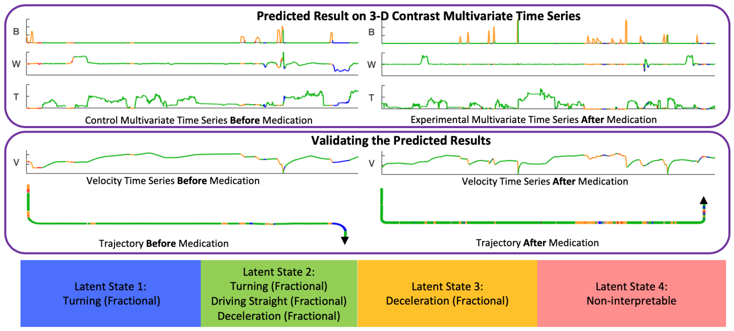


(a) Incorrect latent state assignments.

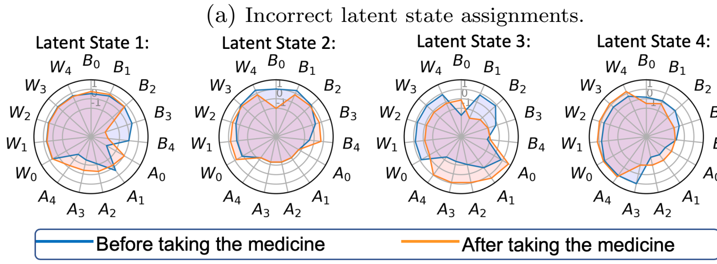


(b) Unreliable contrast patterns.

Figure 2: Applying K-means method on the ADHD medicine controlled experiment



(a) Incorrect latent state assignments.



(b) Unreliable contrast patterns.

Figure 3: Applying K-Shape method on the ADHD medicine controlled experiment

References

[1] L. Erdős, A. Knowles, Yau, et al., Spectral statistics of erdős-rényi graphs i: local semicircle law, The Annals of Probability 41 (3B) (2013) 2279–2375.

[2] D. Hallac, S. Vare, S. Boyd, J. Leskovec, Toeplitz inverse covariance-based clustering of multivariate time series data, in: KDD'17, 2017, pp. 215–223.

[3] K. I. Park, Park, Fundamentals of Probability and Stochastic Processes with Applications to Communications, Springer, 2018.