Latent Sparse Modeling of Longitudinal Multi-dimensional Data

Ko-Shin Chen[1] Tingyang Xu[2] Jinbo Bi[1]

[1]Department of Computer Science and Engineering, University of Connecticut, Storrs, CT, USA

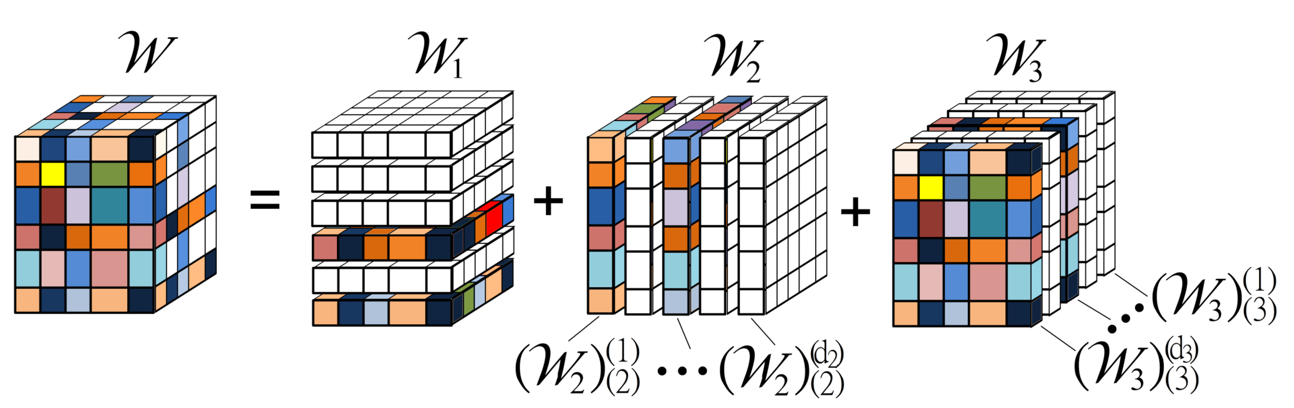
[2]Tencent AI Lab, Shenzhen, China

Introduction

In this work, we propose a tensor-based model where outcomes not only depend on the current observation but also from multiple previous consecutive observations. Simultaneously the model determines the temporal contingency and the most influential features along each dimension of the tensor data.

K-way Tensor & latent $L_{F,1}$ Regularizer

- We decompose a K-way tensor W as $W = \sum_{k=1}^{K} W_k$ such that W_k is sparse along the k-th direction.
- Example ($K = 3$)



- To select the most important time points and features, we introduce the latent $L_{F,1}$ regularizer:

$$R(\Phi) = \sum_{k=1}^{K} \lambda_k \left(\sum_{j=1}^{d_k} \left\| (W_k)_{(k)}^{(j)} \right\|_F \right)$$

- Here $\Phi = (W_1, \dots, W_K)$ and $\|\cdot\|_F$ denotes the Frobenius norm of a tensor.

Model Formulation

- Assume the outcome y_t depends on current and previous τ observations: $\chi_t, \chi_{t-1}, \chi_{t-2}, \dots, \chi_{t-\tau}$. The covariate $\chi_{(t)} = [\chi_t, \chi_{t-1}, \chi_{t-2}, \dots, \chi_{t-\tau}]$ is a K-way tensor.
- We apply the *quadratic inference function* (QIF) to deal with within sample correlation so that the correlation structure does not need to be pre-specified.

$$\min_{\Phi=(W_1,\dots,W_K)} F(\Phi) \coloneqq Q(\chi, y - \mathbb{E}[y], W) + R(\Phi)$$

such that $W = \sum_{k=1}^{K} W_k$.

Algorithm: Linearized Block Coordinate Descent

Given ${W_1}^{(r)}, \dots, {W_K}^{(r)}$, for each fixed k, we obtain the updated ${W_k}^{(r+1)}$ by solving

$$\min_{W_k} \frac{1}{2} \left\| W_k - \left({W_k}^{(r)} - C \nabla Q^{(r)}\right) \right\|_F^2 + C\lambda_k \left(\sum_{j=1}^{d_k} \left\| (W_k)_{(k)}^{(j)} \right\|_F \right)$$

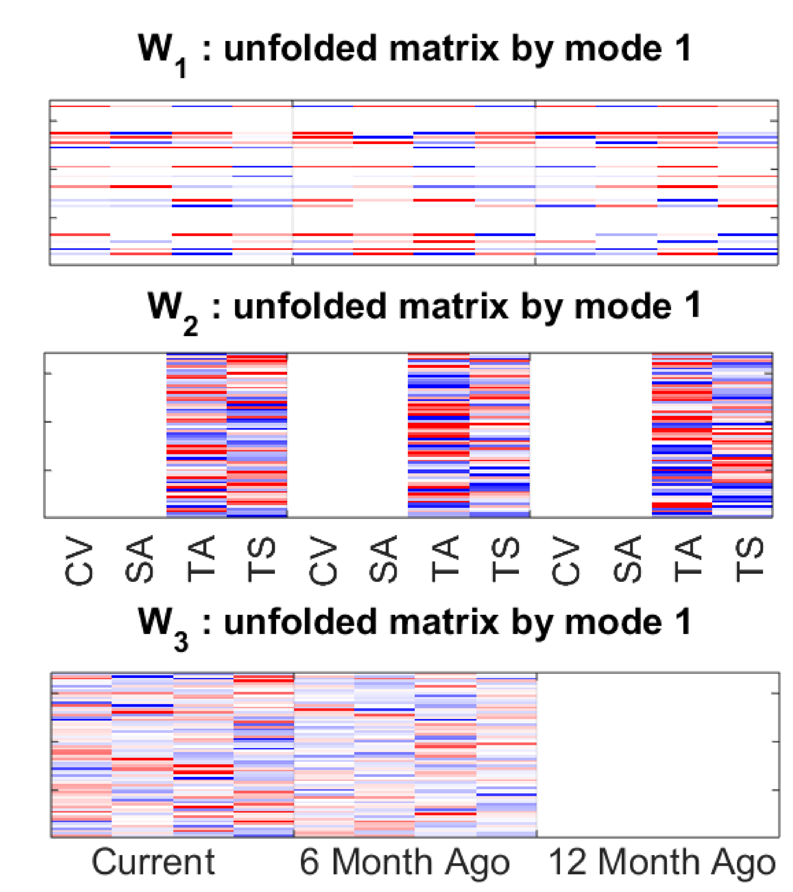
Theorems

- **Theorem 1**: Under certain assumptions, when the sample size $m \to \infty$, we have $\widehat{W} \to W^*$ in probability and $\sqrt{m}\,\mathrm{vect}\left(\widehat{W} - W^*\right) \to \mathcal{N}(0, \Sigma)$ in distribution for some Σ.
- **Theorem 2**: If the initial tuple $\Phi^{(0)}$ is within a convex neighborhood of a minimizer $\widehat{\Phi}$, the algorithm converges and $F\left(\Phi^{(r)}\right) - F\left(\widehat{\Phi}\right) \le \frac{\sum \left\| W_k^{(0)} - \widehat{W} \right\|_F^2}{2rC}$ for all $r \ge 1$.

Feature Selection

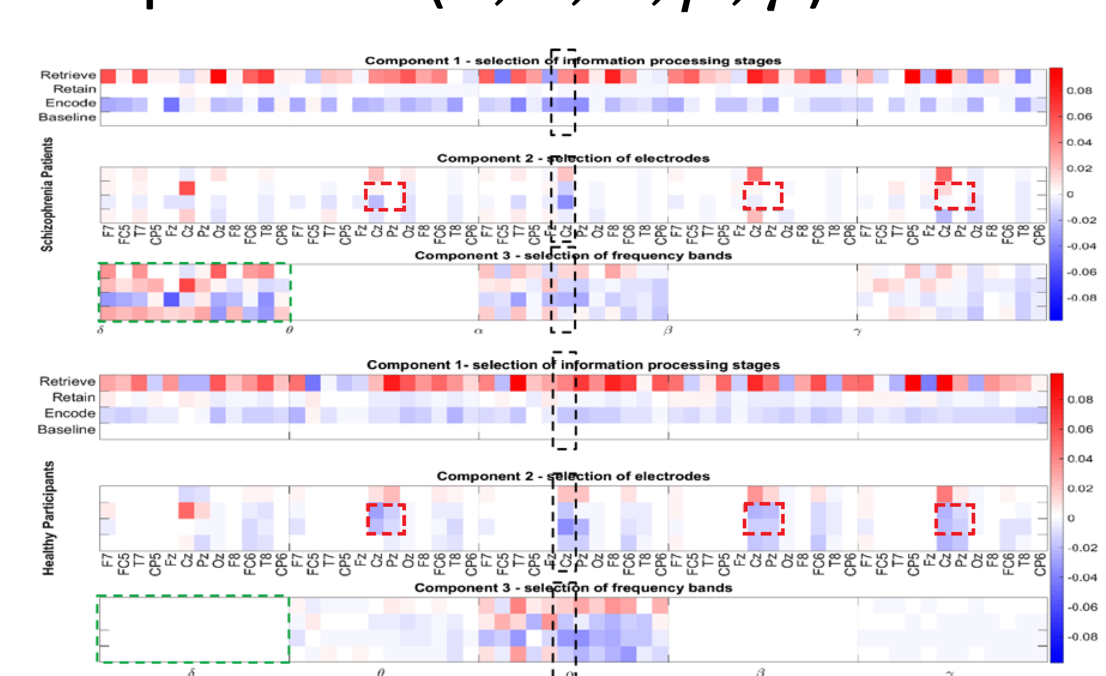
fMRI Data

- Goal: to predict MMSE score.
- Direction 1: brain areas (67).
- Direction 2: brain properties (4).
  - CV – Cortical Volume
  - SA – Surface Area
  - TA – Thickness Average
  - TS – Thickness SD
- Direction 3: time (month).



EEG Data

- Goal: to predict responses (± 1) on Sternberg task.
- Direction 1: processing stages
  (baseline, encoding, retention, retrieval)
- Direction 2: electrodes (12).
- Direction 3: frequency components ($\delta, \theta, \alpha, \beta, \gamma$).
- Features selected for schizophrenia patients and healthy control participants.



Conclusion

1. The *tensorQIF* model finds true coefficient.
2. The corresponding optimization problem can be solved by an algorithm with a sublinear convergence rate.
3. The proposed method enabled interpretation and summary across all dimensions.

Acknowledgements: This work was partially supported by NSF grants DBI-1356655, CCF-1514357, IIS-1718738, as well as NIH grants R01DA037349 and K02DA043063 to Jinbo Bi.