IEEE BIBM 2022

6–8 December • Las Vegas, NV, USA & Changsha, China

Tensor-Based Multi-Modal Multi-Target Regression for Alzheimer's Disease Prediction

Jun Yu, Benjamin Zalatan, Yong Chen, Li Shen, and Lifang He

Speaker: Jun Yu

Date: December 6, 2022

LEHIGH UNIVERSITY

Penn UNIVERSITY of PENNSYLVANIA

Outline

- Background
- Related Work
- Method
- Experiment
- Conclusion

Outline

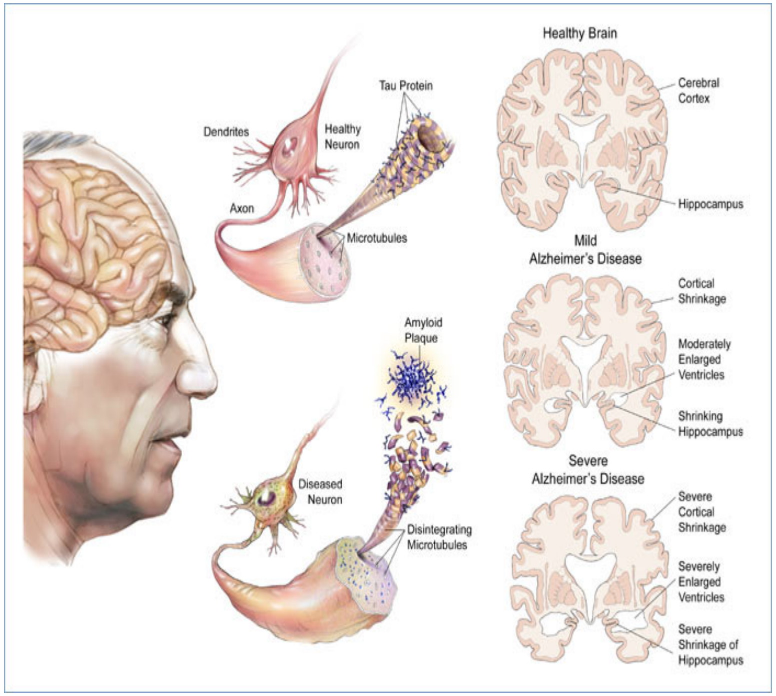
- Background
- Related Work
- Method
- Experiment
- Conclusion

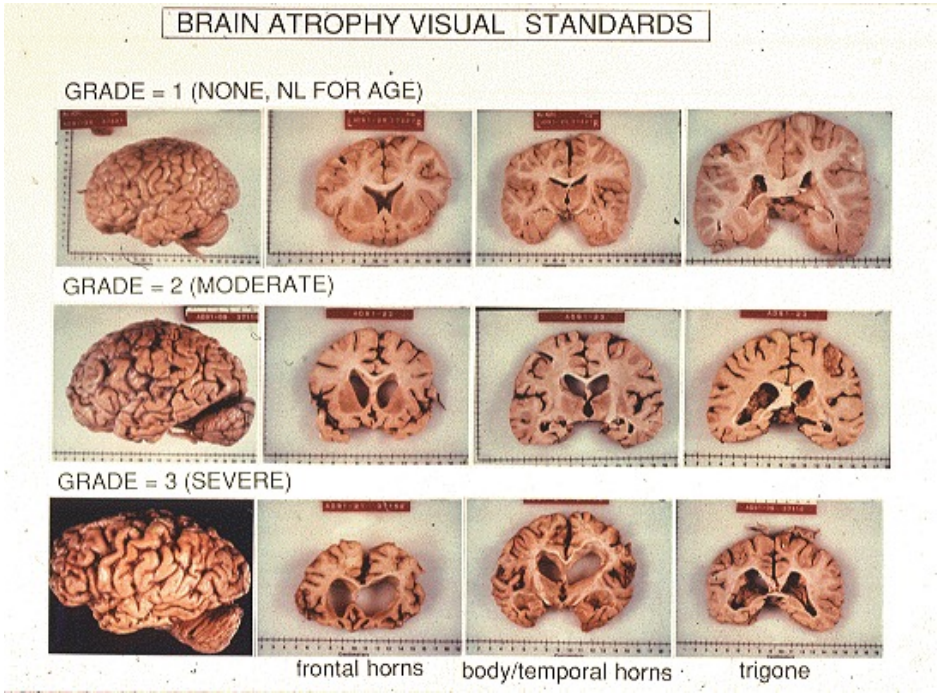
What is Alzheimer’s Disease?

Alzheimer’s Disease (AD) is one of the most common and incurable neurodegenerative diseases, which can result in progressive cognitive decline and behavioral impairment, and even cause death in severe cases.

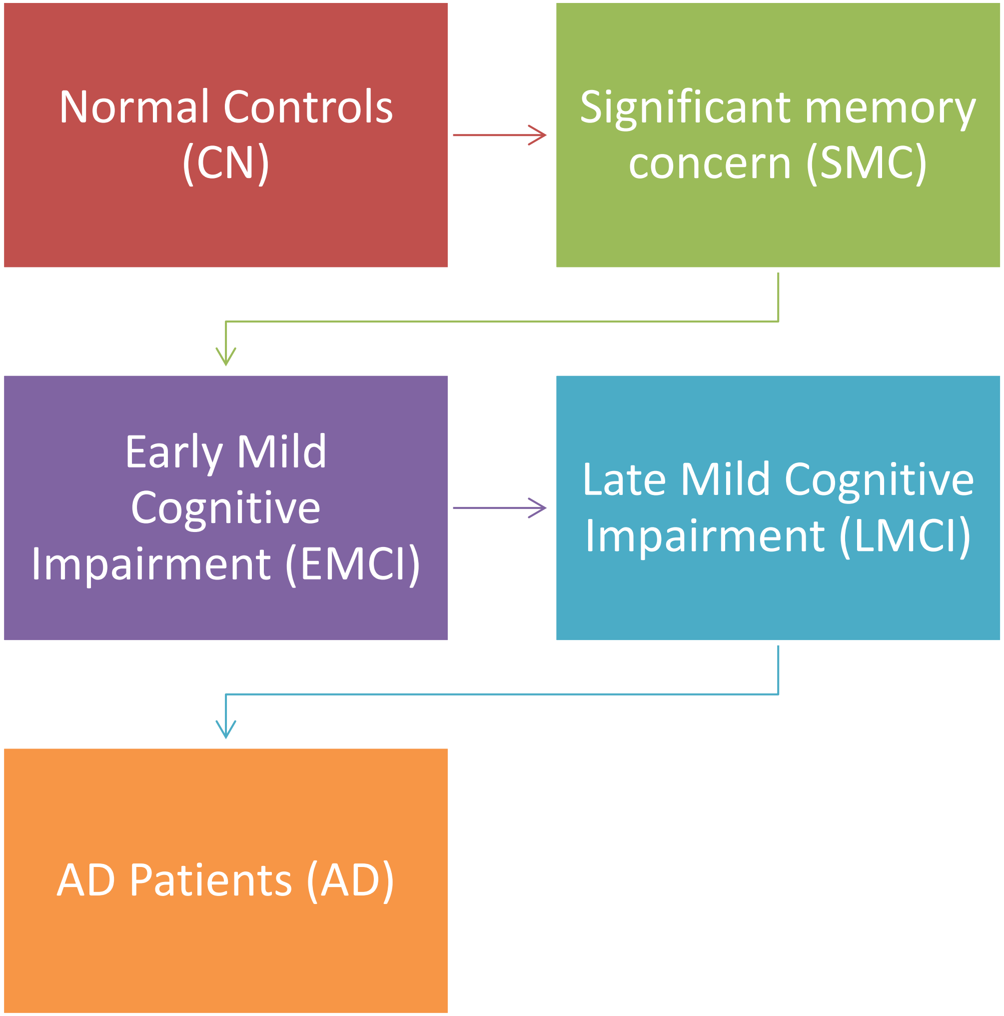
Physical Changes in Brain

- Degradation from cell to organ:





Five Stages of AD



Normal Controls (CN)

Significant memory concern (SMC)

Early Mild Cognitive Impairment (EMCI)

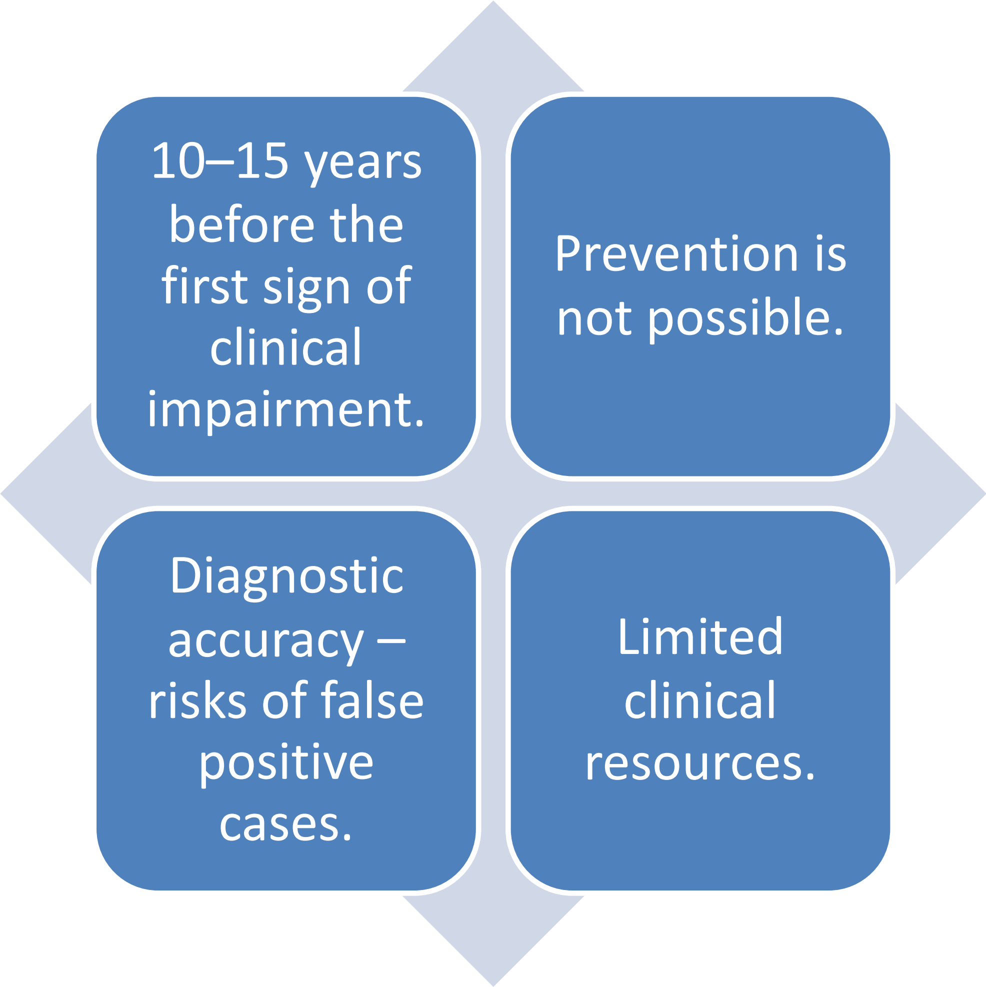
Late Mild Cognitive Impairment (LMCI)

AD Patients (AD)

Statistics of AD in U.S.

- 5,000,000+ detected.
- 20,000,000+ affected.
- 1 AD developed per minute.
- 6th cause of death (4.23%).
- 1st cause of dementia among people age 65+.
- $100,000,000+ caring cost per year.

Challenges in Diagnosis



10–15 years before the first sign of clinical impairment.

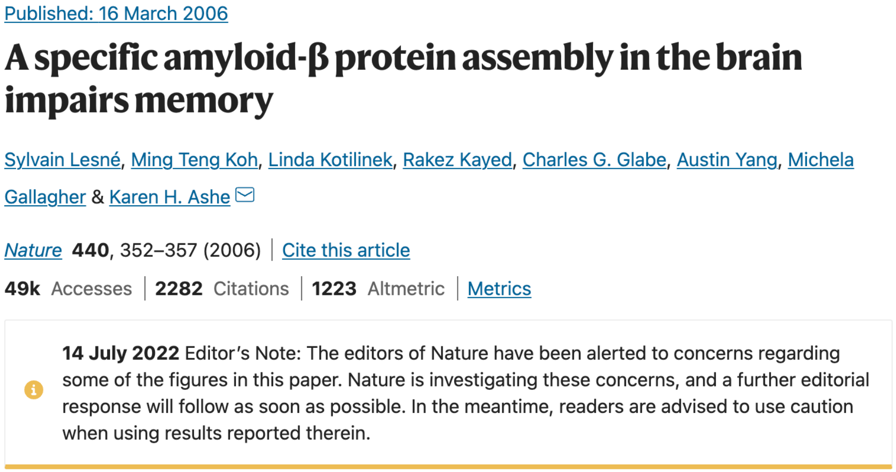
Prevention is not possible.

Diagnostic accuracy – risks of false positive cases.

Limited clinical resources.

Causes of AD

- No one fully understands AD.
- Possible causes: genetic, environmental, and lifestyle factors.
- Aggregation of amyloid-β protein leading to neuroinflammation (possibly fake research!!!)



Published: 16 March 2006

A specific amyloid-β protein assembly in the brain impairs memory

Sylvain Lesné, Ming Teng Koh, Linda Kotilinek, Rakez Kayed, Charles G. Glabe, Austin Yang, Michela Gallagher & Karen H. Ashe

Nature **440**, 352–357 (2006) | Cite this article

49k Accesses | **2282** Citations | **1223** Altmetric | Metrics

14 July 2022 Editor's Note: The editors of Nature have been alerted to concerns regarding some of the figures in this paper. Nature is investigating these concerns, and a further editorial response will follow as soon as possible. In the meantime, readers are advised to use caution when using results reported therein.

Outline

- Background
- Related Work
- Method
- Experiment
- Conclusion

Existing Methods

Feature selection methods

- Vector-based
- Single-modality

PCA-based method

- Feature projection
- Less interpretability

Advanced models (CNNs, GCNs)

- High accuracy
- No interpretability

Baselines

- Sparse MTR models
  - Sparse Multi-Task Regression and Feature selection (SMART): $\ell_{2,1}$ norm
  - Multi-Task Sparse Group Lasso (MT-SGL): Group Lasso
  - Robust Multi-Label Transfer Feature Learning (rMLTFL): $\ell_{2,1}$ norm
- Advanced Models
  - Deep Belief Network-based Multi-Task Learning (DBN-based MTL)
  - Graph Convolutional Neural Network (GCN)

Materials

692 non-Hispanic Caucasian participants in the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database

- 163 CN
- 73 SMC
- 214 EMCI
- 149 LMCI
- 93 AD patients

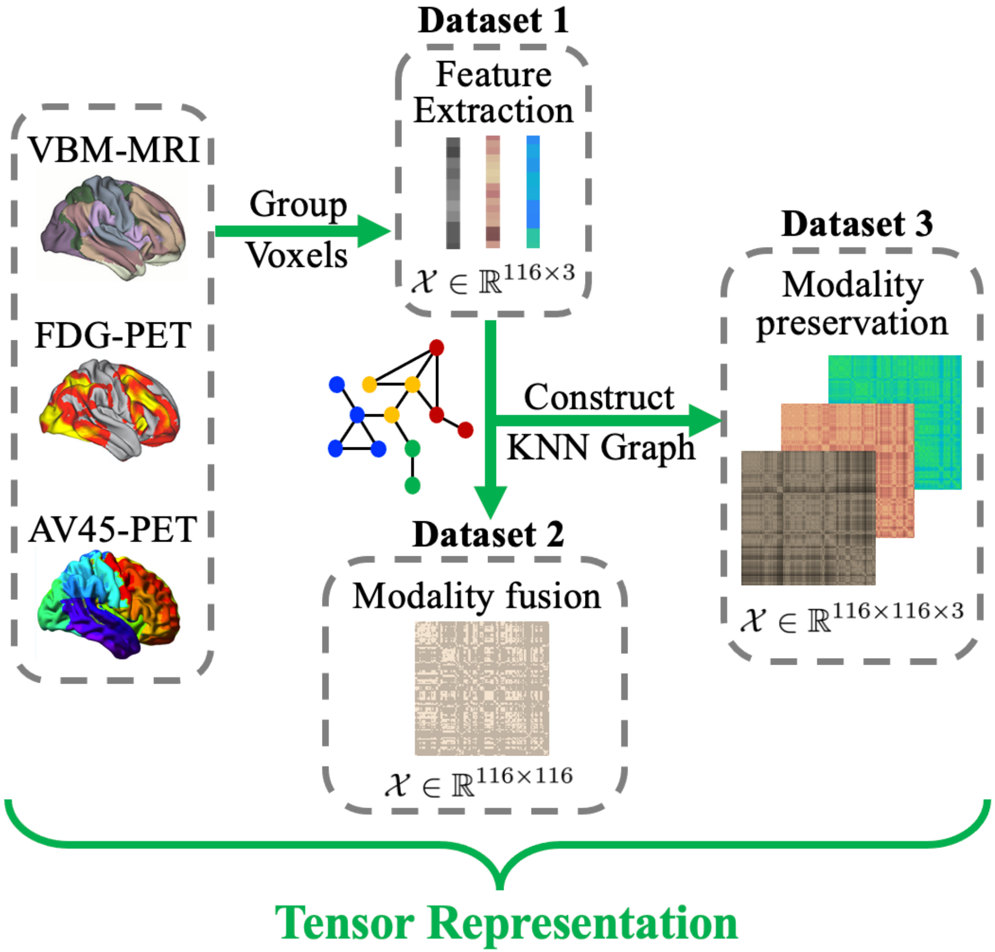
3 modalities of imaging data

- structural Magnetic Resonance Imaging (VBM-MRI)
- 18 F-fluorodeoxyglucose Positron Emission Tomography (FDG-PET)
- 18 F-florbetapir PET (AV45-PET)

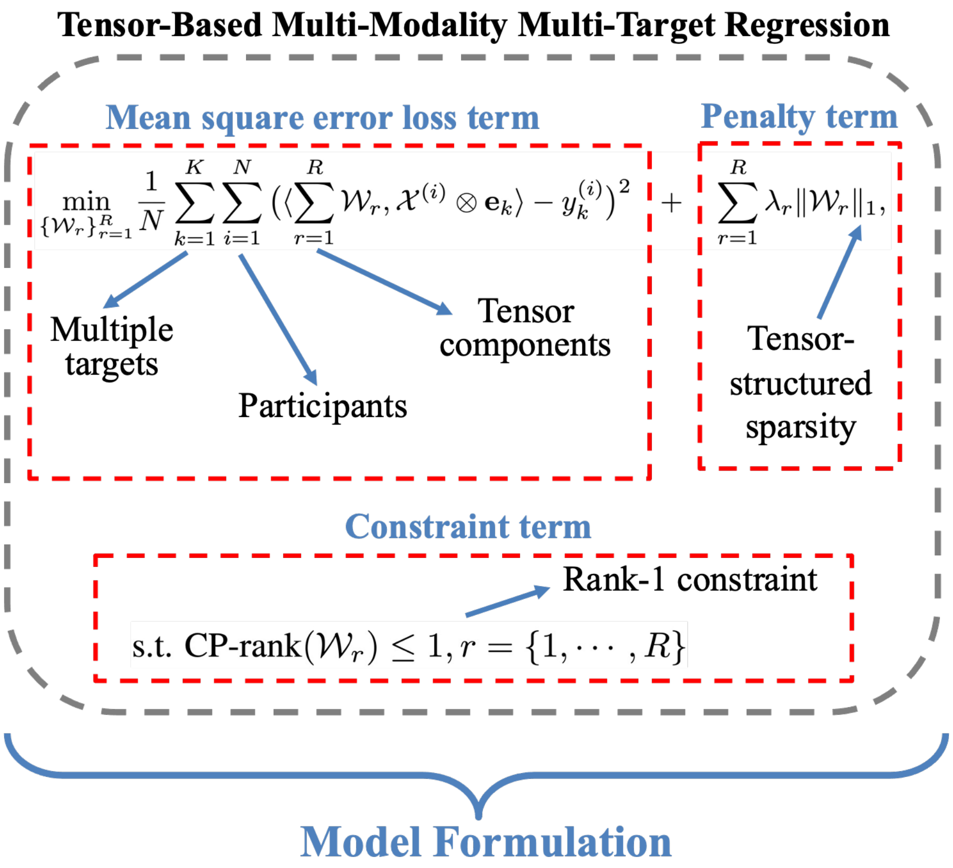
Outline

- Background
- Related Work
- Method
- Experiment
- Conclusion

Data Preprocessing



Formulation



Tensor-Based Multi-Modality Multi-Target Regression

Mean square error loss term

Penalty term

$$\min_{\{\mathcal{W}_r\}_{r=1}^R} \frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}\Big(\langle\sum_{r=1}^{R}\mathcal{W}_r, \mathcal{X}^{(i)}\otimes \mathbf{e}_k\rangle - y_k^{(i)}\Big)^2 + \sum_{r=1}^{R}\lambda_r\|\mathcal{W}_r\|_1,$$

Multiple targets

Participants

Tensor components

Tensor-structured sparsity

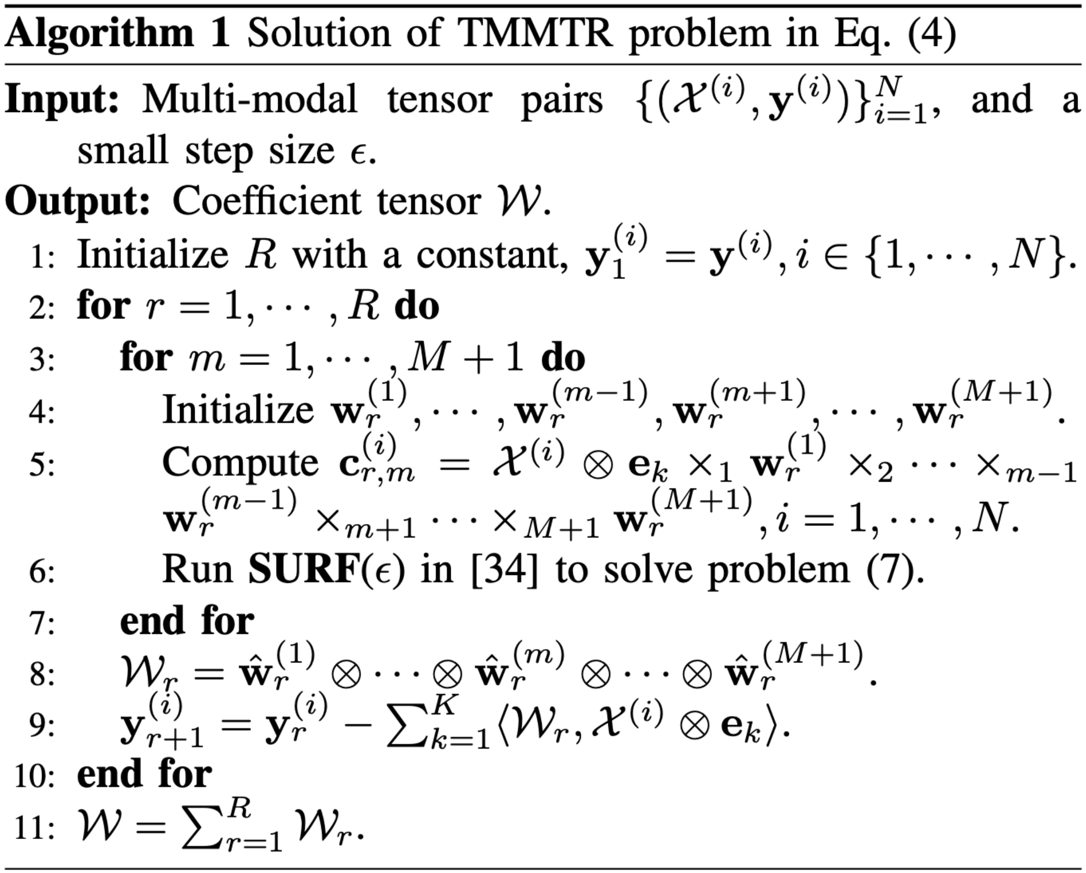
Constraint term

Rank-1 constraint

s.t. CP-rank$(\mathcal{W}_r) \leq 1, r = \{1, \cdots, R\}$

Model Formulation

Algorithm



Algorithm 1 Solution of TMMTR problem in Eq. (4)

```
Input: Multi-modal tensor pairs $\{(\mathcal{X}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^{N}$, and a small step size $\epsilon$.
Output: Coefficient tensor $\mathcal{W}$.
Initialize $R$ with a constant, $\mathbf{y}_1^{(i)} = \mathbf{y}^{(i)}, i \in \{1, \cdots, N\}$.
for $r = 1, \cdots, R$ do
    for $m = 1, \cdots, M+1$ do
        Initialize $\mathbf{w}_r^{(1)}, \cdots, \mathbf{w}_r^{(m-1)}, \mathbf{w}_r^{(m+1)}, \cdots, \mathbf{w}_r^{(M+1)}$.
        Compute $\mathbf{c}_{r,m}^{(i)} = \mathcal{X}^{(i)} \otimes \mathbf{e}_k \times_1 \mathbf{w}_r^{(1)} \times_2 \cdots \times_{m-1} \mathbf{w}_r^{(m-1)} \times_{m+1} \cdots \times_{M+1} \mathbf{w}_r^{(M+1)}, i = 1, \cdots, N$.
        Run SURF($\epsilon$) in [34] to solve problem (7).
    end for
    $\mathcal{W}_r = \hat{\mathbf{w}}_r^{(1)} \otimes \cdots \otimes \hat{\mathbf{w}}_r^{(m)} \otimes \cdots \otimes \hat{\mathbf{w}}_r^{(M+1)}$.
    $\mathbf{y}_{r+1}^{(i)} = \mathbf{y}_r^{(i)} - \sum_{k=1}^{K} \langle \mathcal{W}_r, \mathcal{X}^{(i)} \otimes \mathbf{e}_k \rangle$.
end for
$\mathcal{W} = \sum_{r=1}^{R} \mathcal{W}_r$.
```

Outline

- Background
- Related Work
- Method
- **Experiment**
- Conclusion

Performance Comparison

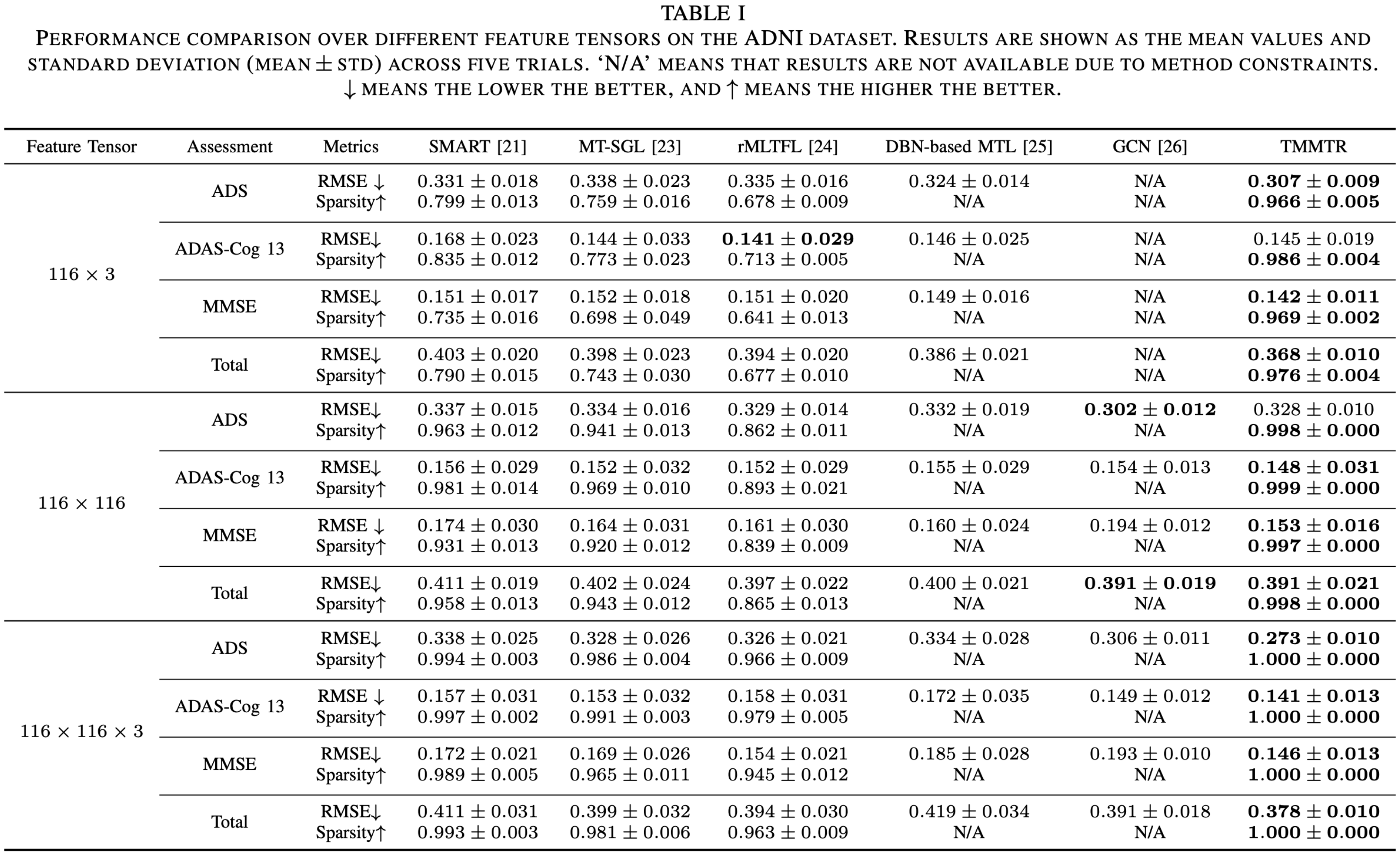
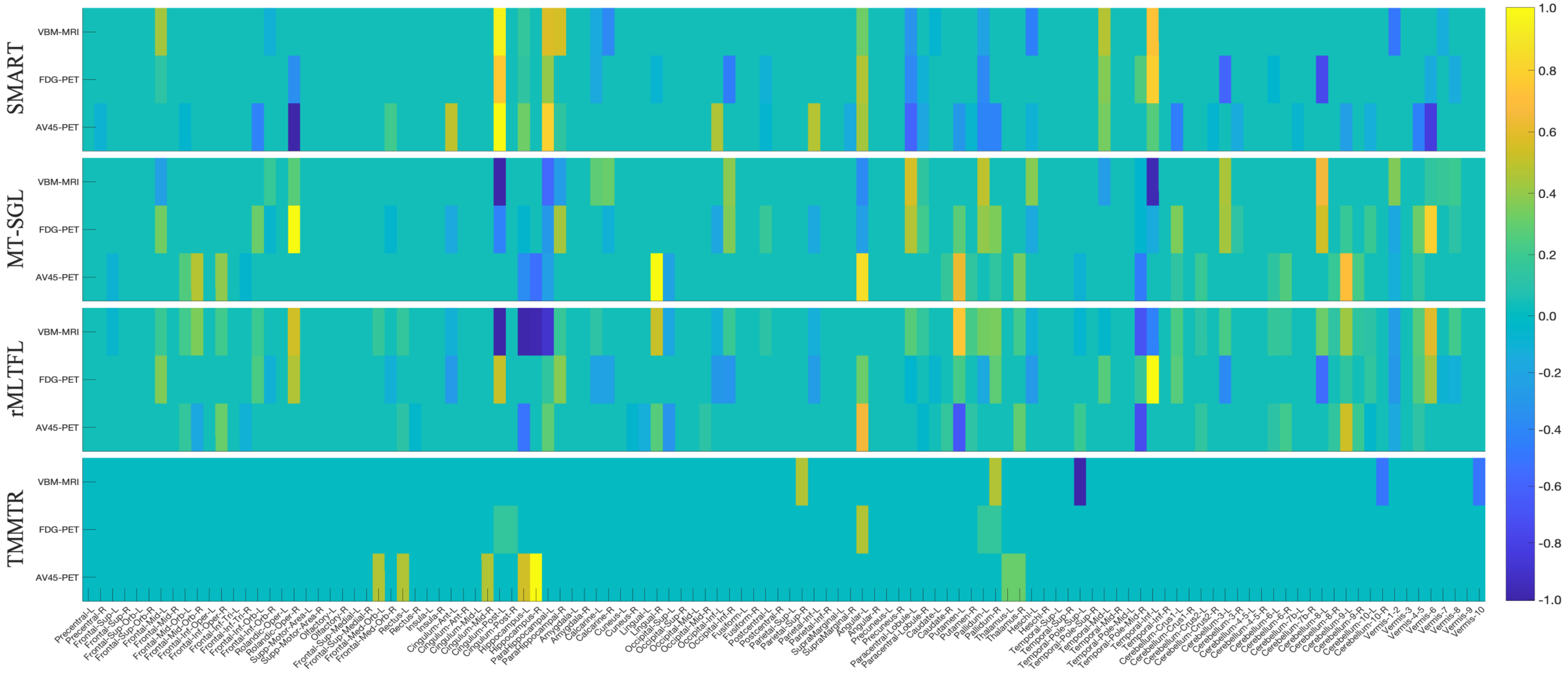


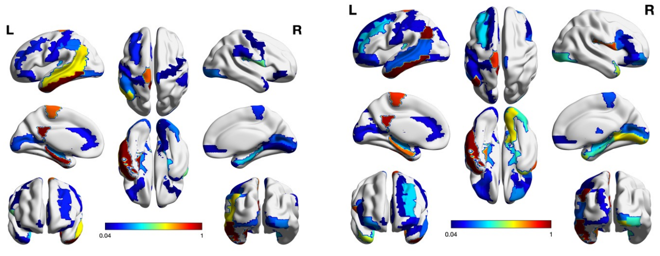
TABLE I

Performance comparison over different feature tensors on the ADNI dataset. Results are shown as the mean values and standard deviation (mean ± std) across five trials. 'N/A' means that results are not available due to method constraints. ↓ means the lower the better, and ↑ means the higher the better.

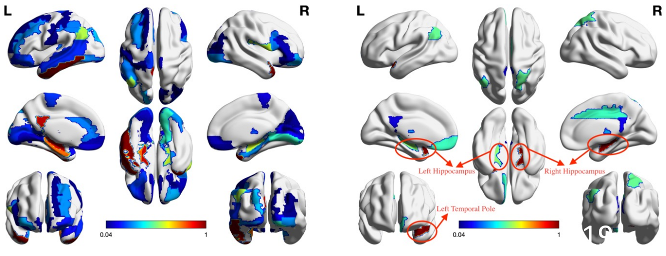
Feature Tensor	Assessment	Metrics	SMART [21]	MT-SGL [23]	rMLTFL [24]	DBN-based MTL [25]	GCN [26]	TMMTR
116×3	ADS	RMSE ↓	0.331 ± 0.018	0.338 ± 0.023	0.335 ± 0.016	0.324 ± 0.014	N/A	$\mathbf{0.307 \pm 0.009}$
		Sparsity↑	0.799 ± 0.013	0.759 ± 0.016	0.678 ± 0.009	N/A	N/A	$\mathbf{0.966 \pm 0.005}$
	ADAS-Cog 13	RMSE↓	0.168 ± 0.023	0.144 ± 0.033	$\mathbf{0.141 \pm 0.029}$	0.146 ± 0.025	N/A	0.145 ± 0.019
		Sparsity↑	0.835 ± 0.012	0.773 ± 0.023	0.713 ± 0.005	N/A	N/A	$\mathbf{0.986 \pm 0.004}$
	MMSE	RMSE↓	0.151 ± 0.017	0.152 ± 0.018	0.151 ± 0.020	0.149 ± 0.016	N/A	$\mathbf{0.142 \pm 0.011}$
		Sparsity↑	0.735 ± 0.016	0.698 ± 0.049	0.641 ± 0.013	N/A	N/A	$\mathbf{0.969 \pm 0.002}$
	Total	RMSE↓	0.403 ± 0.020	0.398 ± 0.023	0.394 ± 0.020	0.386 ± 0.021	N/A	$\mathbf{0.368 \pm 0.010}$
		Sparsity↑	0.790 ± 0.015	0.743 ± 0.030	0.677 ± 0.010	N/A	N/A	$\mathbf{0.976 \pm 0.004}$
116×116	ADS	RMSE↓	0.337 ± 0.015	0.334 ± 0.016	0.329 ± 0.014	0.332 ± 0.019	$\mathbf{0.302 \pm 0.012}$	0.328 ± 0.010
		Sparsity↑	0.963 ± 0.012	0.941 ± 0.013	0.862 ± 0.011	N/A	N/A	$\mathbf{0.998 \pm 0.000}$
	ADAS-Cog 13	RMSE↓	0.156 ± 0.029	0.152 ± 0.032	0.152 ± 0.029	0.155 ± 0.029	0.154 ± 0.013	$\mathbf{0.148 \pm 0.031}$
		Sparsity↑	0.981 ± 0.014	0.969 ± 0.010	0.893 ± 0.021	N/A	N/A	$\mathbf{0.999 \pm 0.000}$
	MMSE	RMSE ↓	0.174 ± 0.030	0.164 ± 0.031	0.161 ± 0.030	0.160 ± 0.024	0.194 ± 0.012	$\mathbf{0.153 \pm 0.016}$
		Sparsity↑	0.931 ± 0.013	0.920 ± 0.012	0.839 ± 0.009	N/A	N/A	$\mathbf{0.997 \pm 0.000}$
	Total	RMSE↓	0.411 ± 0.019	0.402 ± 0.024	0.397 ± 0.022	0.400 ± 0.021	$\mathbf{0.391 \pm 0.019}$	$\mathbf{0.391 \pm 0.021}$
		Sparsity↑	0.958 ± 0.013	0.943 ± 0.012	0.865 ± 0.013	N/A	N/A	$\mathbf{0.998 \pm 0.000}$
$116 \times 116 \times 3$	ADS	RMSE↓	0.338 ± 0.025	0.328 ± 0.026	0.326 ± 0.021	0.334 ± 0.028	0.306 ± 0.011	$\mathbf{0.273 \pm 0.010}$
		Sparsity↑	0.994 ± 0.003	0.986 ± 0.004	0.966 ± 0.009	N/A	N/A	$\mathbf{1.000 \pm 0.000}$
	ADAS-Cog 13	RMSE ↓	0.157 ± 0.031	0.153 ± 0.032	0.158 ± 0.031	0.172 ± 0.035	0.149 ± 0.012	$\mathbf{0.141 \pm 0.013}$
		Sparsity↑	0.997 ± 0.002	0.991 ± 0.003	0.979 ± 0.005	N/A	N/A	$\mathbf{1.000 \pm 0.000}$
	MMSE	RMSE↓	0.172 ± 0.021	0.169 ± 0.026	0.154 ± 0.021	0.185 ± 0.028	0.193 ± 0.010	$\mathbf{0.146 \pm 0.013}$
		Sparsity↑	0.989 ± 0.005	0.965 ± 0.011	0.945 ± 0.012	N/A	N/A	$\mathbf{1.000 \pm 0.000}$
	Total	RMSE↓	0.411 ± 0.031	0.399 ± 0.032	0.394 ± 0.030	0.419 ± 0.034	0.391 ± 0.018	$\mathbf{0.378 \pm 0.010}$
		Sparsity↑	0.993 ± 0.003	0.981 ± 0.006	0.963 ± 0.009	N/A	N/A	$\mathbf{1.000 \pm 0.000}$

Visualization





(a) SMART (b) MT-SGL



(c) rMLTFL (d) TMMTR

Ablation Study

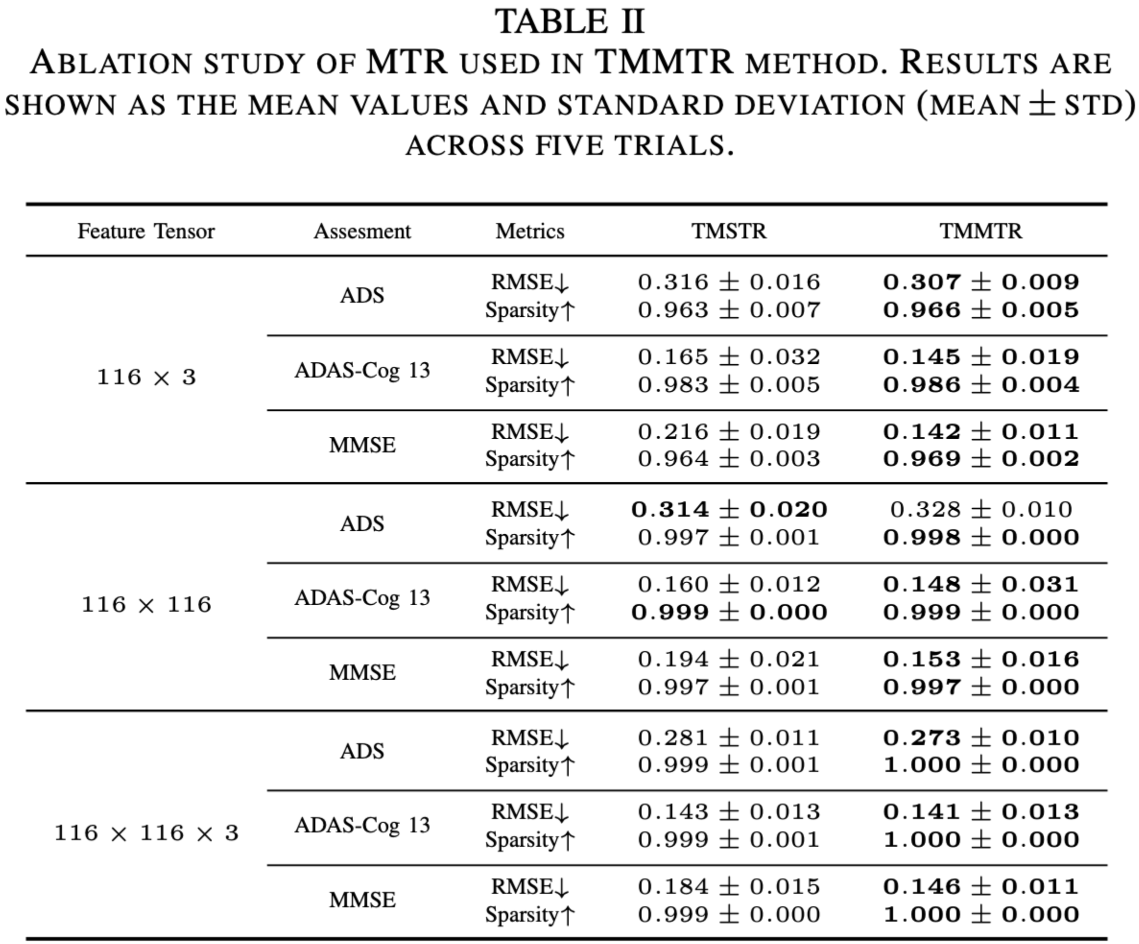
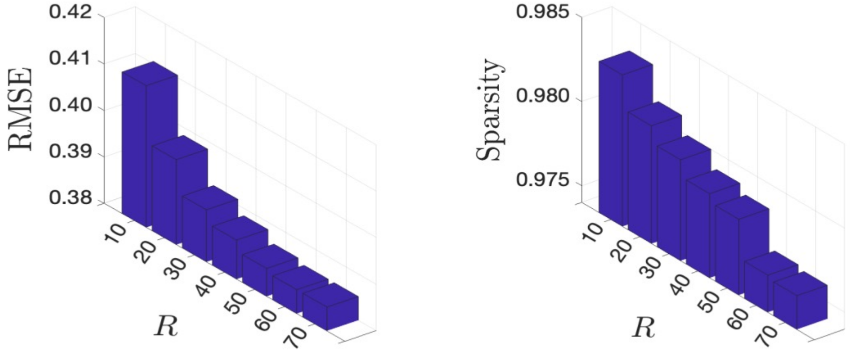


TABLE II

Ablation study of MTR used in TMMTR method. Results are shown as the mean values and standard deviation (mean ± std) across five trials.

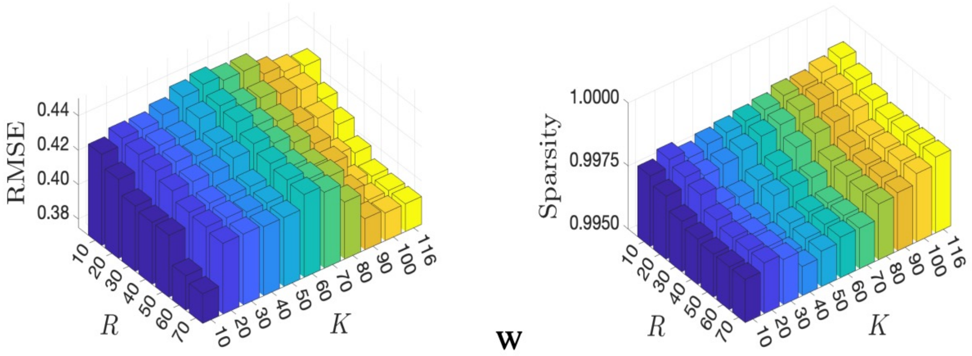
Feature Tensor	Assesment	Metrics	TMSTR	TMMTR
116×3	ADS	RMSE↓ Sparsity↑	0.316 ± 0.016 0.963 ± 0.007	$\mathbf{0.307 \pm 0.009}$ $\mathbf{0.966 \pm 0.005}$
	ADAS-Cog 13	RMSE↓ Sparsity↑	0.165 ± 0.032 0.983 ± 0.005	$\mathbf{0.145 \pm 0.019}$ $\mathbf{0.986 \pm 0.004}$
	MMSE	RMSE↓ Sparsity↑	0.216 ± 0.019 0.964 ± 0.003	$\mathbf{0.142 \pm 0.011}$ $\mathbf{0.969 \pm 0.002}$
116×116	ADS	RMSE↓ Sparsity↑	$\mathbf{0.314 \pm 0.020}$ 0.997 ± 0.001	0.328 ± 0.010 $\mathbf{0.998 \pm 0.000}$
	ADAS-Cog 13	RMSE↓ Sparsity↑	0.160 ± 0.012 $\mathbf{0.999 \pm 0.000}$	$\mathbf{0.148 \pm 0.031}$ $\mathbf{0.999 \pm 0.000}$
	MMSE	RMSE↓ Sparsity↑	0.194 ± 0.021 0.997 ± 0.001	$\mathbf{0.153 \pm 0.016}$ $\mathbf{0.997 \pm 0.000}$
$116 \times 116 \times 3$	ADS	RMSE↓ Sparsity↑	0.281 ± 0.011 0.999 ± 0.001	$\mathbf{0.273 \pm 0.010}$ $\mathbf{1.000 \pm 0.000}$
	ADAS-Cog 13	RMSE↓ Sparsity↑	0.143 ± 0.013 0.999 ± 0.001	$\mathbf{0.141 \pm 0.013}$ $\mathbf{1.000 \pm 0.000}$
	MMSE	RMSE↓ Sparsity↑	0.184 ± 0.015 0.999 ± 0.000	$\mathbf{0.146 \pm 0.011}$ $\mathbf{1.000 \pm 0.000}$

Hyperparameter Analysis



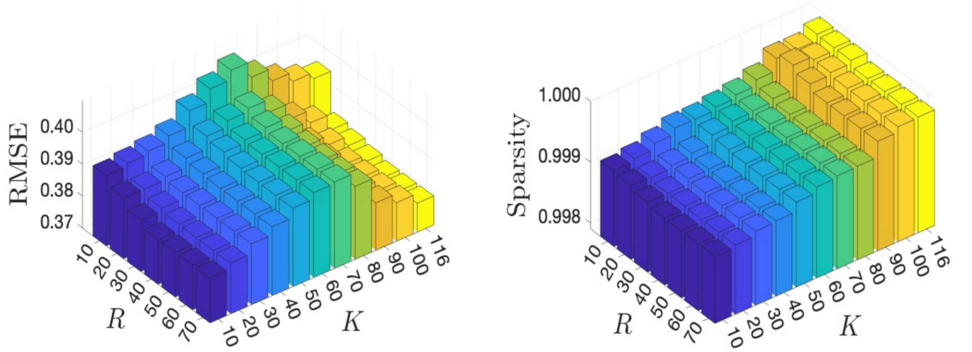
(a) RMSE of 116 × 3 dataset

(b) Sparsity of 116 × 3 dataset



(c) RMSE of 116 × 116 dataset

(d) Sparsity of 116 × 116 dataset



(e) RMSE of 116 × 116 × 3 dataset

(f) Sparsity of 116 × 116 × 3 dataset

Outline

- Background
- Related Work
- Method
- Experiment
- Conclusion

Conclusion

Tensor-structured information and Inter-target correlation are leveraged in TMMTR.

The Divide-and-conquer algorithm to solve TMMTR is effective and efficient.

Better performance with higher sparsity is realized in TMMTR.