

Nonparametric Factor Trajectory Learning for Dynamic Tensor Decomposition

Zheng Wang, Shandian Zhe

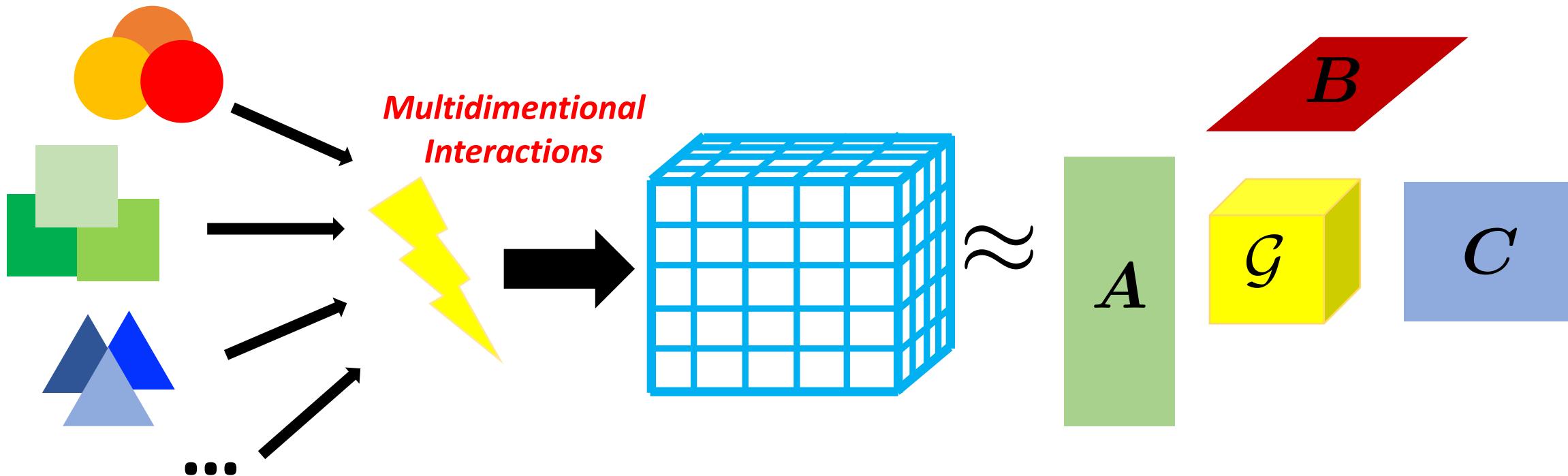
School of Computing

University of Utah



SCHOOL OF COMPUTING
UNIVERSITY OF UTAH

Quick Overview

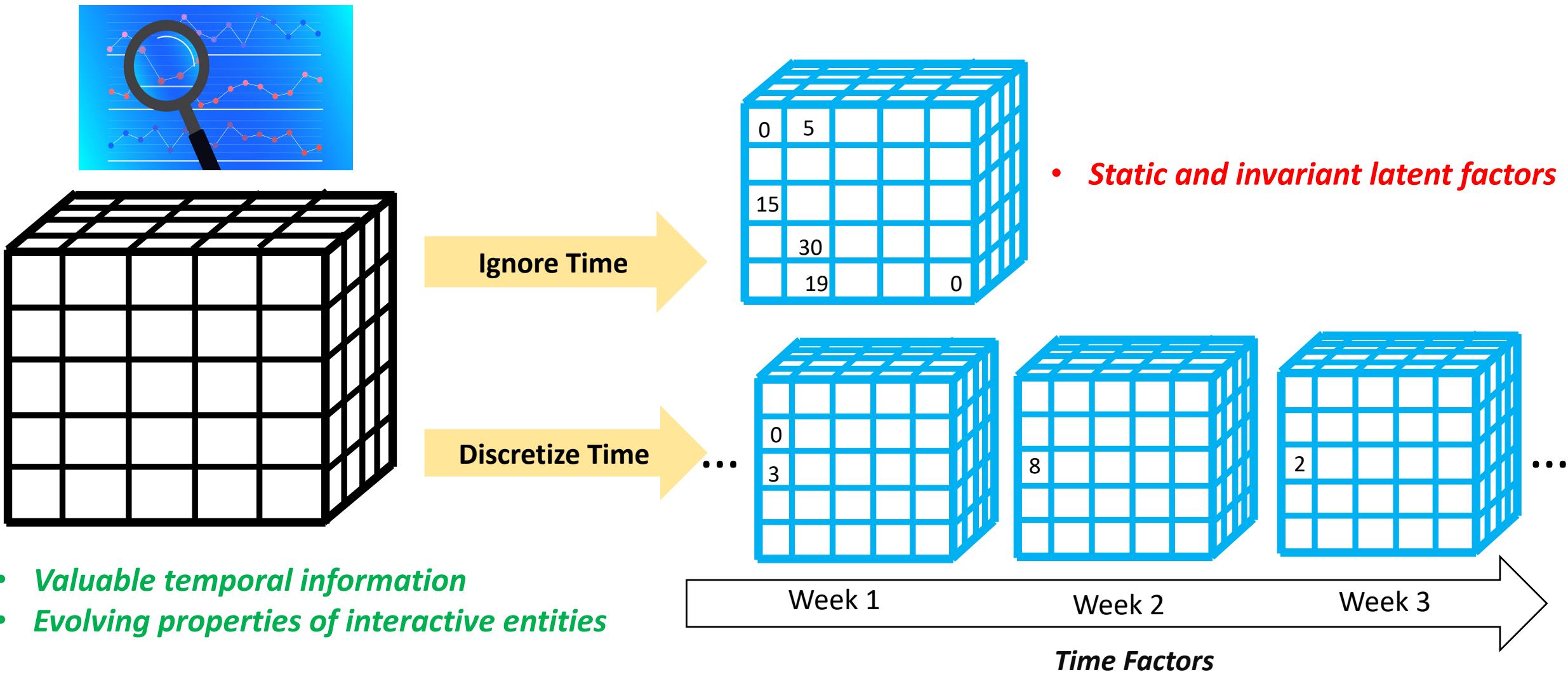


- Customer, Store, Product
- Location, Age, Gender
- Website, Location, Ads Type
- Latitude, Longitude, Elevation
- ...

- ✓ Online Store Transactions
- ✓ Social Media User Behaviors
- ✓ Advertisement Click Log Data
- ✓ Global Climate Data
- ✓ ...

- Tucker decomposition (Tucker, 1966)
- CANDECOMP/PARAFAC (CP) decomposition (Harshman, 1970)
- ...

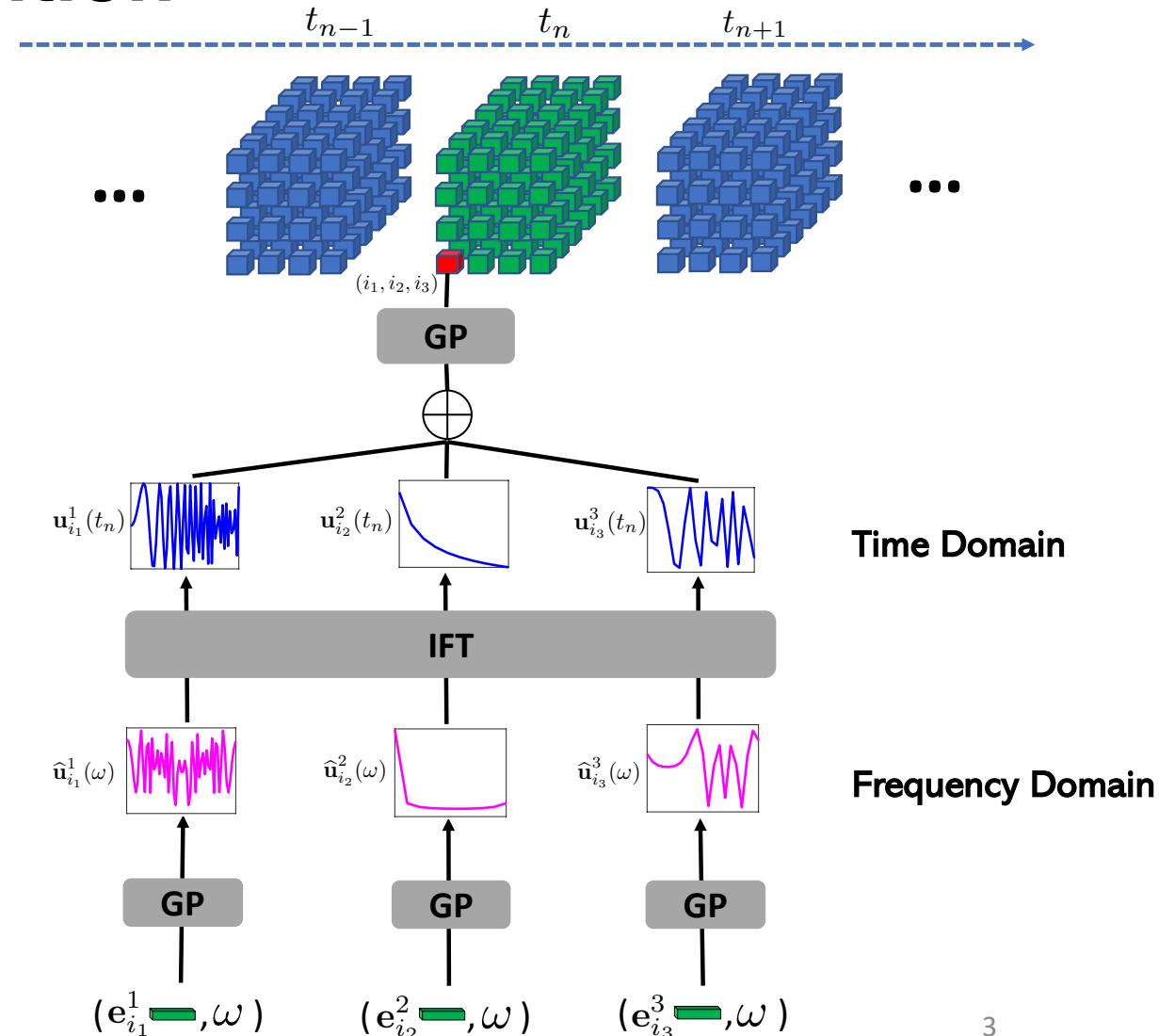
Underexplored Dynamic Tensor Factorization



Nonparametric Factor Trajectory Learning for Dynamic Tensor Decomposition

- **Major Contributions:**

- First nonlinear decomposition method for factor trajectory learning
- Long-term trajectory estimation
- Robust, flexible for sparse, noisy data



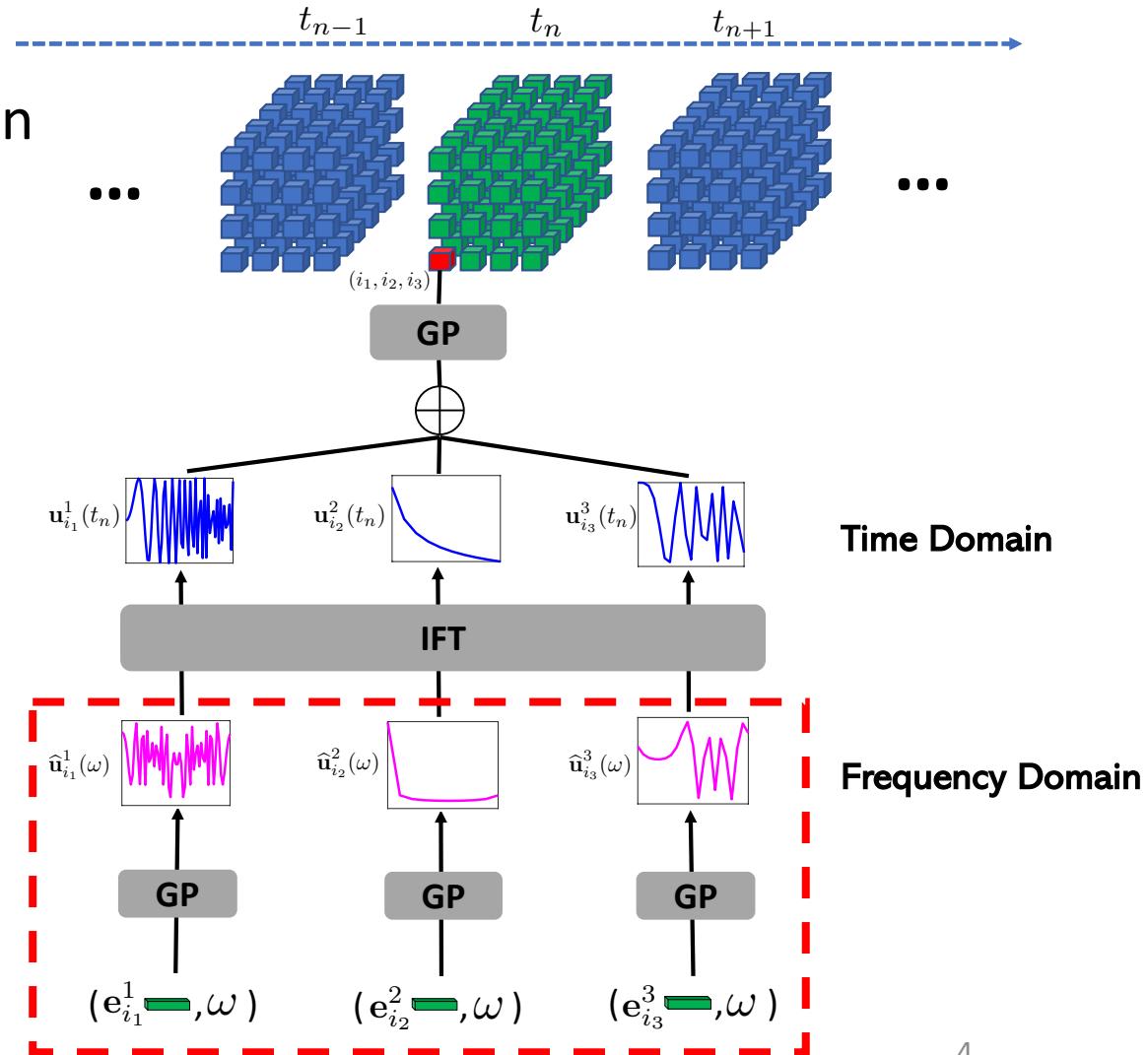
Spectral Tensor Decomposition – First-Layer GP

- Frequency embeddings are learned based on entity embeddings and frequency values

$$\mathbf{E}^k = [\mathbf{e}_1^k, \dots, \mathbf{e}_{d_k}^k]^\top$$

$$\hat{\boldsymbol{\omega}} = [\hat{\omega}_1; \dots; \hat{\omega}_C]$$

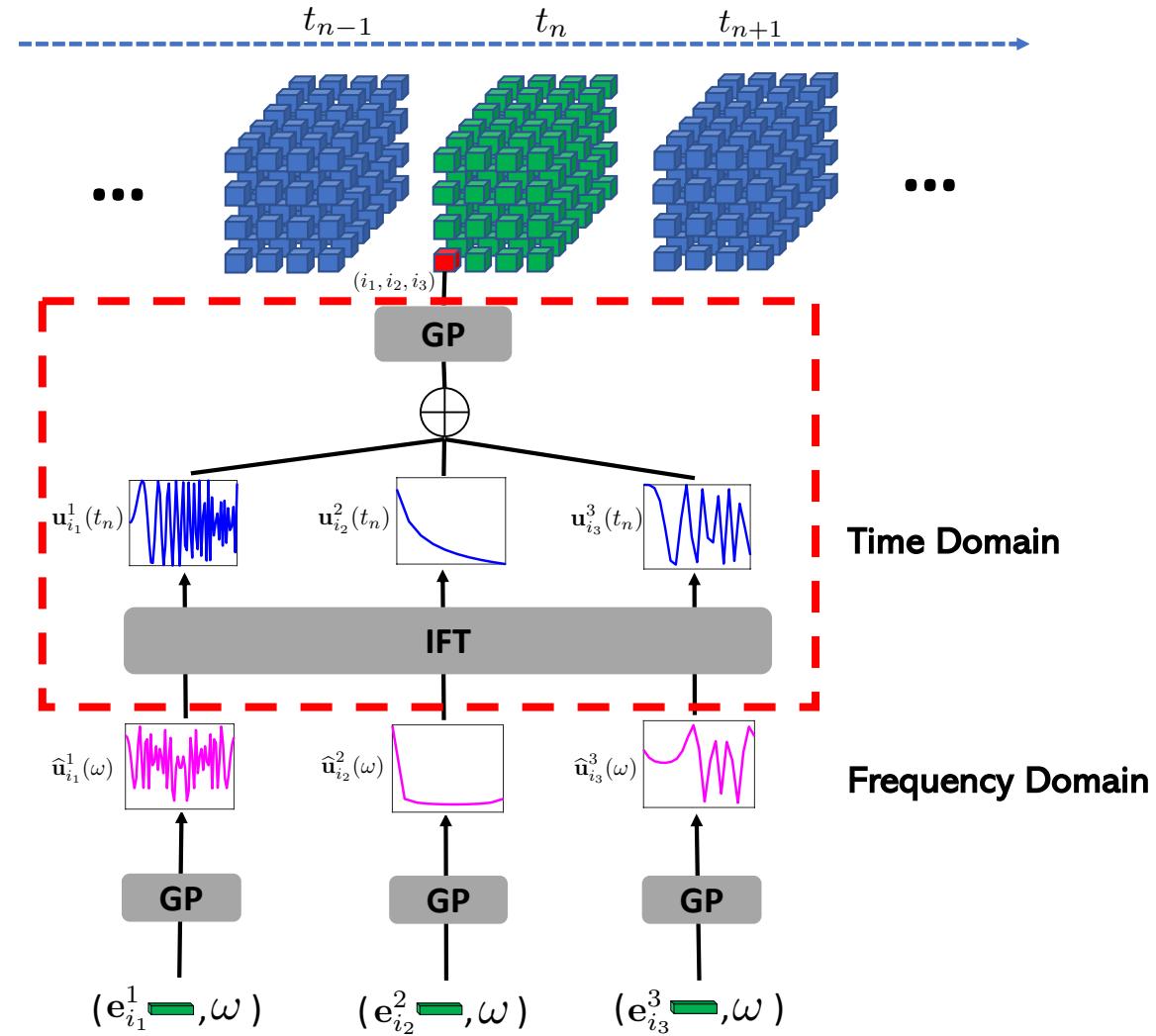
$$\begin{aligned} p(\mathbf{F}_r^k) &= N(\mathbf{f}_r^k | \mathbf{0}, \kappa_r(\mathbf{E}^k, \mathbf{E}^k) \otimes \kappa_r(\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\omega}})) \\ &= \mathcal{MN}(\mathbf{F}_r^k | \mathbf{0}, \kappa_r(\mathbf{E}^k, \mathbf{E}^k), \kappa_r(\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\omega}})) \\ &= \frac{\exp\left(-\frac{1}{2}\text{tr}\left(\kappa_r(\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\omega}})^{-1} (\mathbf{F}_r^k)^\top \kappa_r(\mathbf{E}^k, \mathbf{E}^k)^{-1} \mathbf{F}_r^k\right)\right)}{(2\pi)^{d_k C/2} |\kappa_r(\mathbf{E}^k, \mathbf{E}^k)|^{C/2} |\kappa_r(\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\omega}})|^{d_k/2}}. \end{aligned} \quad (15)$$



Temporal Tensor Decomposition – Second-Layer GP

- Temporal trajectories are recovered by applying inverse Fourier transform (IFT)
- Nonlinear tensor factorization by applying GP

$$m_{\ell}(t) = g(\mathbf{u}_{\ell_1}^1(t), \dots, \mathbf{u}_{\ell_K}^K(t)) \\ \sim \mathcal{GP}(0, \kappa_g(\mathbf{v}_{\ell}(t), \mathbf{v}_{\ell'}(t))),$$



Nested Stochastic Inference with Sparse GP Prior

- Sparse variational ELBO based on Matrix Gaussian Prior and Posterior

$$p(\{\mathbf{G}_r^k, \mathbf{F}_r^k\}, \mathbf{h}, \mathbf{m}, \mathbf{y}) = \prod_{k=1}^K \prod_{r=1}^R p(\mathbf{G}_r^k) p(\mathbf{F}_r^k | \mathbf{G}_r^k) \cdot p(\mathbf{h}) p(\mathbf{m} | \mathbf{h}) \mathcal{N}(\mathbf{y} | \mathbf{m}, \sigma^2 \mathbf{I}). \quad (19)$$
$$q(\{\mathbf{G}_r^k, \mathbf{F}_r^k\}, \mathbf{h}, \mathbf{m}) = \prod_{k=1}^K \prod_{r=1}^R q(\mathbf{G}_r^k) p(\mathbf{F}_r^k | \mathbf{G}_r^k) q(\mathbf{h}) p(\mathbf{m} | \mathbf{h}). \quad (20)$$

$$\begin{aligned} \mathcal{L} = & - \sum_{k=1}^K \sum_{r=1}^R \text{KL} (q(\mathbf{G}_r^k) \| p(\mathbf{G}_r^k)) \\ & - \text{KL} (q(\mathbf{h}) \| p(\mathbf{h})) + \sum_{n=1}^N \mathbb{E}_q [\log p(y_n | m_{\ell_n}(t_n))], \end{aligned} \quad (21)$$

- Nested Stochastic Mini-Batch Optimization
 - Reparameterization for both GP layers

Experiments – Prediction Error

| <i>Beijing Air Quality</i> | <i>R</i> = 2 | <i>R</i> = 3 | <i>R</i> = 5 | <i>R</i> = 7 |
|----------------------------|----------------------|----------------------|----------------------|----------------------|
| NONFAT | 0.340 ± 0.006 | 0.315 ± 0.001 | 0.314 ± 0.001 | 0.326 ± 0.006 |
| CPCT | 0.997 ± 0.001 | 0.997 ± 0.001 | 1.002 ± 0.002 | 1.002 ± 0.002 |
| GPCT | 0.372 ± 0.001 | 0.366 ± 0.001 | 0.363 ± 0.001 | 0.364 ± 0.001 |
| NNCT | 0.986 ± 0.002 | 0.988 ± 0.002 | 0.977 ± 0.012 | 0.987 ± 0.003 |
| GPDTL | 0.884 ± 0.001 | 0.884 ± 0.001 | 0.885 ± 0.001 | 0.884 ± 0.001 |
| NNDTL | 0.356 ± 0.003 | 0.358 ± 0.005 | 0.333 ± 0.003 | 0.315 ± 0.002 |
| GPDTN | 0.884 ± 0.001 | 0.884 ± 0.001 | 0.884 ± 0.001 | 0.884 ± 0.001 |
| NNDTN | 0.365 ± 0.005 | 0.337 ± 0.006 | 0.336 ± 0.003 | 0.319 ± 0.005 |
| <i>Mobile Ads</i> | | | | |
| NONFAT | 0.652 ± 0.002 | 0.635 ± 0.003 | 0.638 ± 0.006 | 0.637 ± 0.005 |
| CPCT | 1.001 ± 0.004 | 0.986 ± 0.018 | 1.009 ± 0.009 | 0.971 ± 0.010 |
| GPCT | 0.660 ± 0.003 | 0.661 ± 0.003 | 0.662 ± 0.001 | 0.659 ± 0.003 |
| NNCT | 0.822 ± 0.001 | 0.822 ± 0.001 | 0.822 ± 0.001 | 0.822 ± 0.001 |
| GPDTL | 0.714 ± 0.006 | 0.695 ± 0.004 | 0.695 ± 0.004 | 0.695 ± 0.003 |
| NNDTL | 0.646 ± 0.003 | 0.646 ± 0.002 | 0.642 ± 0.003 | 0.640 ± 0.003 |
| GPDTN | 0.667 ± 0.003 | 0.661 ± 0.003 | 0.668 ± 0.003 | 0.669 ± 0.003 |
| NNDTN | 0.646 ± 0.004 | 0.645 ± 0.002 | 0.640 ± 0.003 | 0.638 ± 0.003 |
| <i>DBLP</i> | | | | |
| NONFAT | 0.188 ± 0.003 | 0.188 ± 0.003 | 0.189 ± 0.003 | 0.189 ± 0.003 |
| CPCT | 1.004 ± 0.003 | 1.004 ± 0.002 | 1.005 ± 0.002 | 1.001 ± 0.004 |
| GPCT | 0.189 ± 0.003 | 0.191 ± 0.003 | 0.192 ± 0.003 | 0.196 ± 0.003 |
| NNCT | 0.188 ± 0.003 | 0.188 ± 0.003 | 0.188 ± 0.003 | 0.189 ± 0.003 |
| GPDTL | 0.208 ± 0.004 | 0.223 ± 0.003 | 0.221 ± 0.003 | 0.224 ± 0.003 |
| NNDTL | 0.188 ± 0.003 | 0.188 ± 0.003 | 0.189 ± 0.003 | 0.189 ± 0.003 |
| GPDTN | 0.206 ± 0.002 | 0.218 ± 0.003 | 0.224 ± 0.003 | 0.225 ± 0.002 |
| NNDTN | 0.188 ± 0.003 | 0.188 ± 0.003 | 0.188 ± 0.003 | 0.189 ± 0.003 |

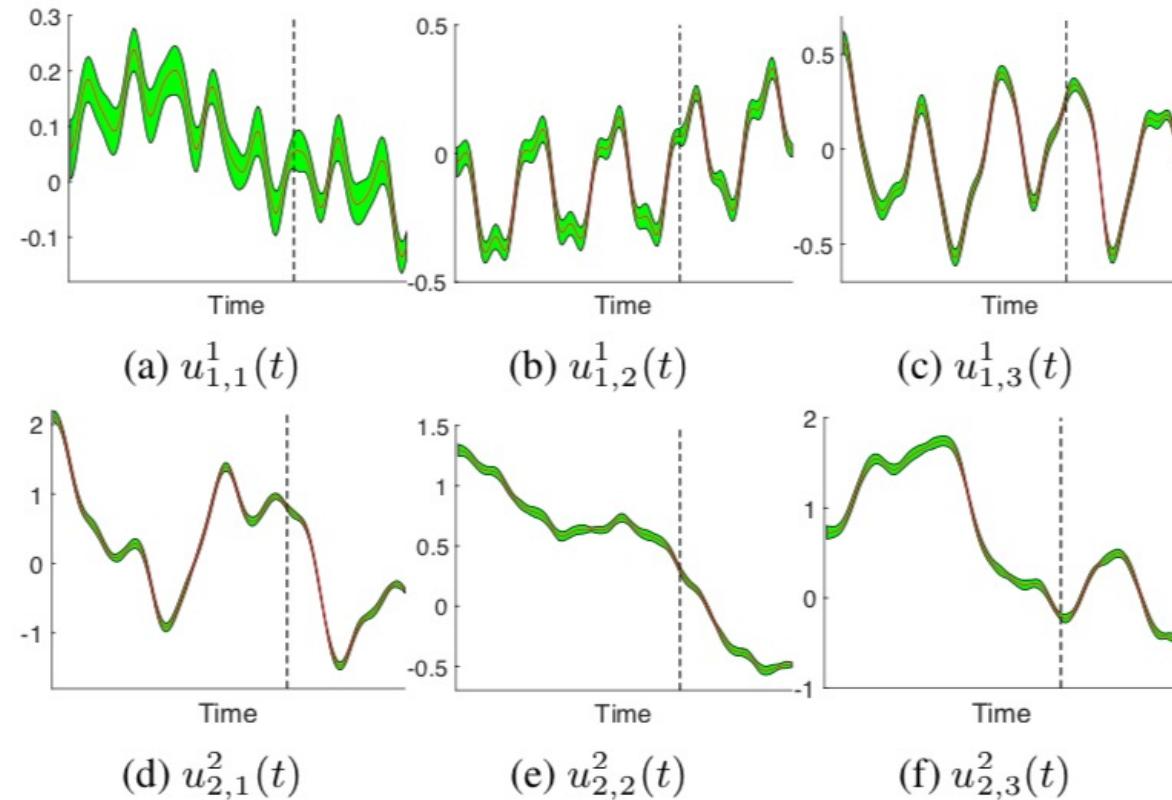
Table 1. Root Mean-Square Error (RMSE). The results were averaged over five runs.

Experiments – Loglikelihood

| <i>Beijing Air Quality</i> | $R = 2$ | $R = 3$ | $R = 5$ | $R = 7$ |
|----------------------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| NONFAT | -0.343 ± 0.018 | -0.264 ± 0.003 | -0.260 ± 0.004 | -0.297 ± 0.017 |
| GPCT | -0.420 ± 0.001 | -0.406 ± 0.001 | -0.401 ± 0.001 | -0.401 ± 0.001 |
| GPDTL | -1.299 ± 0.001 | -1.299 ± 0.001 | -1.299 ± 0.001 | -1.299 ± 0.001 |
| GPDTN | -1.299 ± 0.001 | -1.299 ± 0.001 | -1.299 ± 0.001 | -1.299 ± 0.001 |
| <i>Mobile Ads</i> | | | | |
| NONFAT | -0.726 ± 0.004 | -0.705 ± 0.003 | -0.709 ± 0.007 | -0.706 ± 0.008 |
| GPCT | -0.733 ± 0.002 | -0.737 ± 0.005 | -0.734 ± 0.004 | -0.735 ± 0.004 |
| GPDTL | -1.843 ± 0.009 | -1.807 ± 0.006 | -1.822 ± 0.008 | -1.830 ± 0.003 |
| GPDTN | -0.774 ± 0.003 | -0.762 ± 0.004 | -0.804 ± 0.006 | -0.806 ± 0.003 |
| <i>DBLP</i> | | | | |
| NONFAT | 0.201 ± 0.019 | 0.201 ± 0.019 | 0.199 ± 0.017 | 0.199 ± 0.017 |
| GPCT | 0.129 ± 0.009 | 0.105 ± 0.009 | 0.104 ± 0.011 | 0.087 ± 0.013 |
| GPDTL | 0.102 ± 0.023 | 0.004 ± 0.025 | 0.035 ± 0.019 | 0.022 ± 0.019 |
| GPDTN | 0.114 ± 0.012 | 0.041 ± 0.019 | 0.019 ± 0.020 | 0.013 ± 0.015 |

Table 2. Test log-likelihood. The results were averaged from five runs.

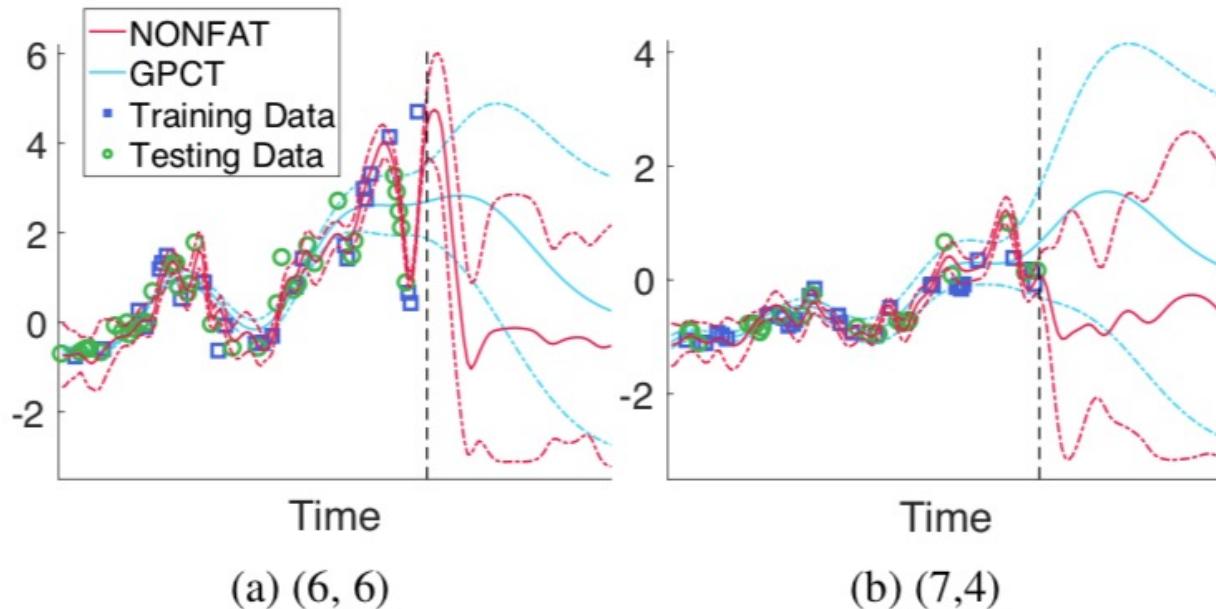
Experiments – Embedding Trajectory Learning



Greatly reduced uncertainties!

Figure 2. The learned factor trajectories.

Experiments – Entry Value Evolution



Better extrapolative uncertainties!

Figure 3. Entry value prediction.

Conclusion

- We have presented NONFAT, a novel nonparametric Bayesian method to *learn factor trajectories for dynamic tensor decomposition*. The predictive accuracy of NONFAT in real-world applications is encouraging and the learned trajectories show interesting temporal patterns.

Thanks!

wzhut@cs.utah.edu
University of Utah