Decomposing Temporal High-Order Interactions via Latent ODEs

Shibo Li, Robert Mike Kirby, Shandian Zhe

School of Computing, Scientific Computing and Imaging Institute; University of Utah

{shibo, kirby, zhe}@cs.utah.edu

Abstract

High-order interactions between multiple objects are common in real-world applications. Although tensor decomposition is a popular framework for high-order interaction analysis and prediction, most methods cannot well exploit the valuable timestamp information in data. The existent methods either discard the timestamps or convert them into discrete steps or use over-simplistic decomposition models. As a result, these methods might not be capable enough of capturing complex, fine-grained temporal dynamics or making accurate predictions for long-term interaction results. To overcome these limitations, we propose a novel Temporal High-order Interaction decomposition model based on Ordinary Differential Equations (THIS-ODE). We model the time-varying interaction result with a latent ODE. To capture the complex temporal dynamics, we use a neural network (NN) to learn the time derivative of the ODE state. We use the representation of the interaction objects to model the initial value of the ODE and to constitute a part of the NN input to compute the state. In this way, the temporal relationships of the participant objects can be estimated and encoded into their representations. For tractable and scalable inference, we use forward sensitivity analysis to efficiently compute the gradient of ODE state, based on which we use integral transform to develop a stochastic mini-batch learning algorithm. We demonstrate the advantage of our approach in simulation and four real-world applications.

Introduction

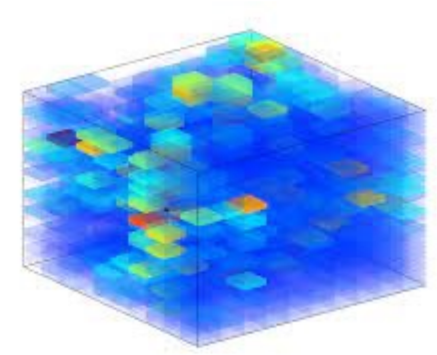
High-order *interactions* in real-world:

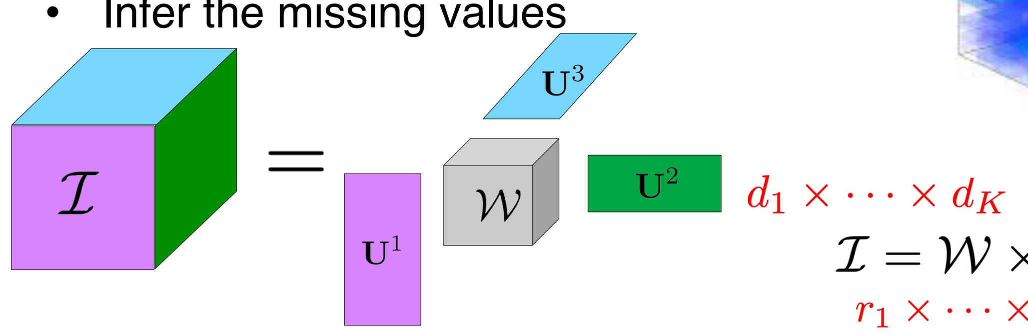
- *Customers* purchase *items* at different *grocery stores*
- *People* take outdoor *exercises* at various *places*

(user, movie, episode) (user, advertisement, page-section) (user, user, location, #hashtag)

Conventional Tensor Decomposition:

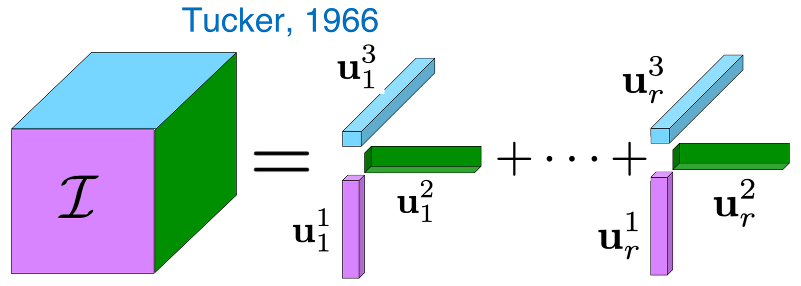
- Internal structure
- Compact representation
- Infer the missing values





$\mathcal{I} = \mathcal{W} \times_1 \mathbf{U}^1 \times_2 \cdots \times_K \mathbf{U}^K$

Tucker, 1966

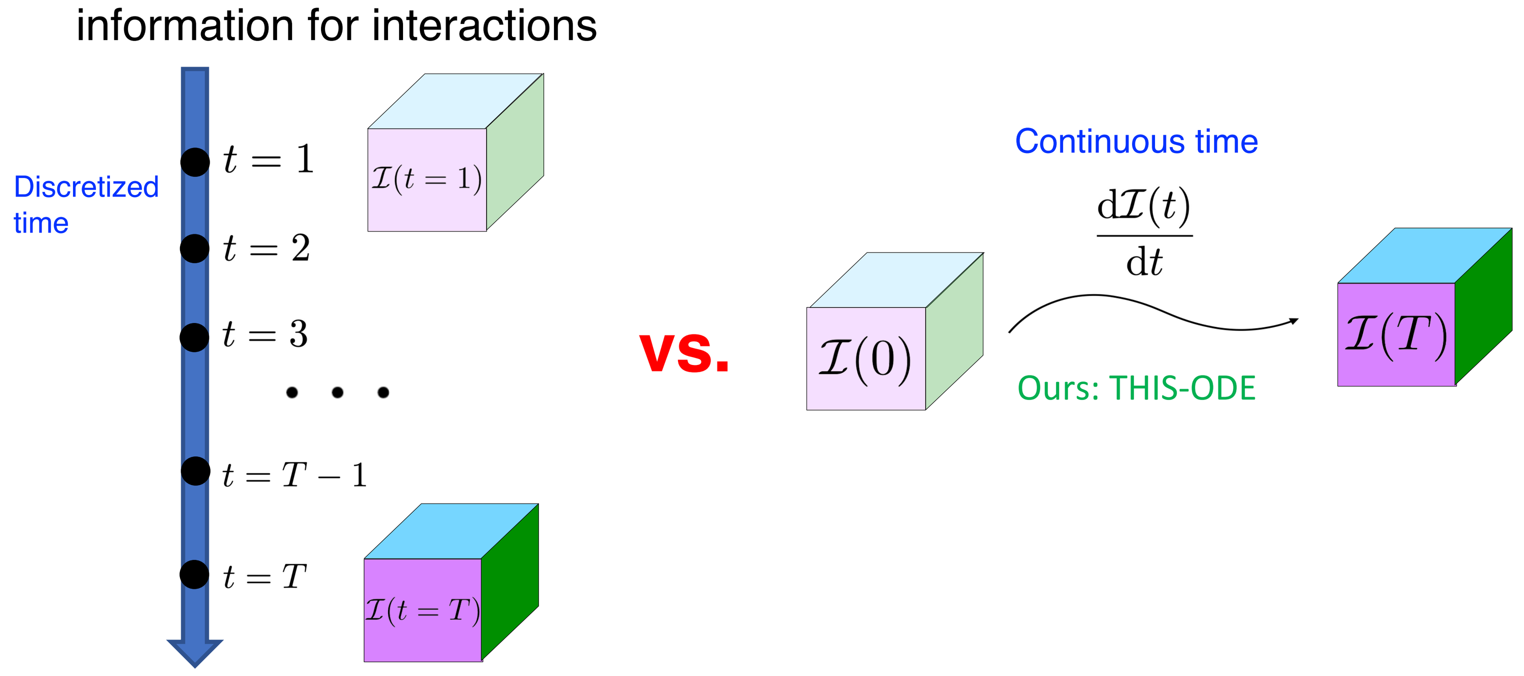


$\mathcal{I} = \sum_{j=1}^{r} \lambda_j \cdot \mathbf{U}^1[:, j] \circ \ldots \circ \mathbf{U}^K[:, j]$

CANDECOMP/PARAFAC (CP), Harshman, 1970

Temporal Interactions : $\mathcal{I}(t)$

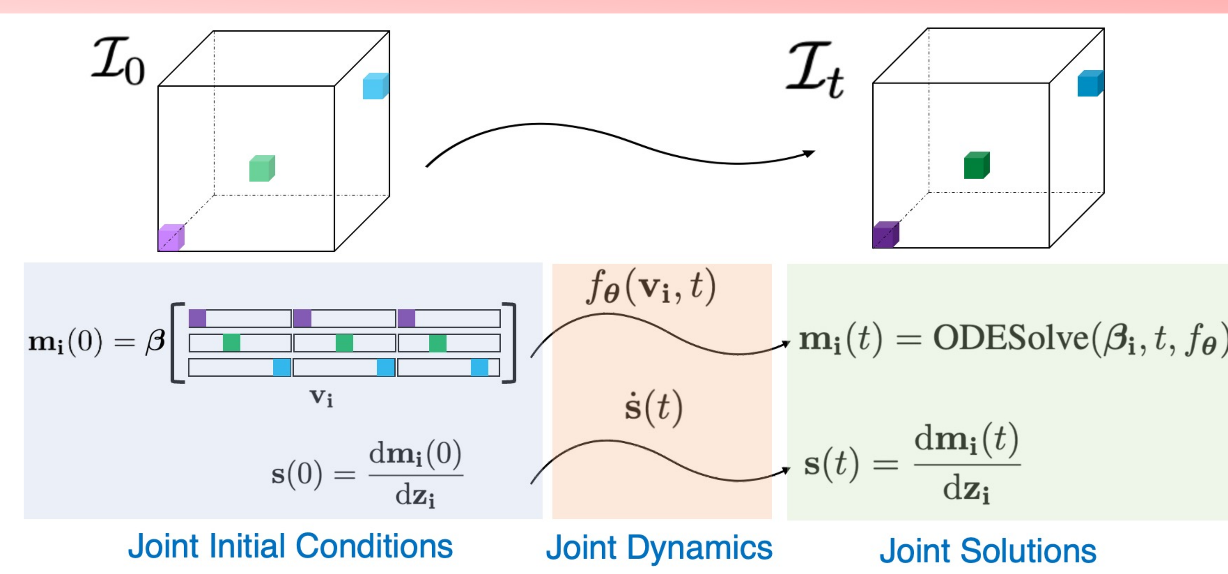
- Interactions are functions of time (complex temporal dynamics)
- Current methods: Discretize the timestamps or simply ignore temporal information for interactions



Our Contribution:

- ***THIS-ODE***: A novel decomposing model of temporal high-order interactions.
- Leverage continuous timestamps, capture all kinds of complex temporal dynamics within interactions.
- Tractable inference with forward sensitivity analysis and time alignment/integral transform tricks.

Methods



Temporal interaction as parametric ODE

$$\begin{cases} \frac{d m_i(t)}{dt} = f(m_i(t), \mathbf{v}_i, t) \\ m_i(0) = \beta(\mathbf{v}_i) \end{cases} \xrightarrow{\text{ODESolve}} m_i(t) = m_i(0) + \int_0^t f_\theta(m_i(s), \mathbf{v}_i, s) ds$$

Given by ODE solvers

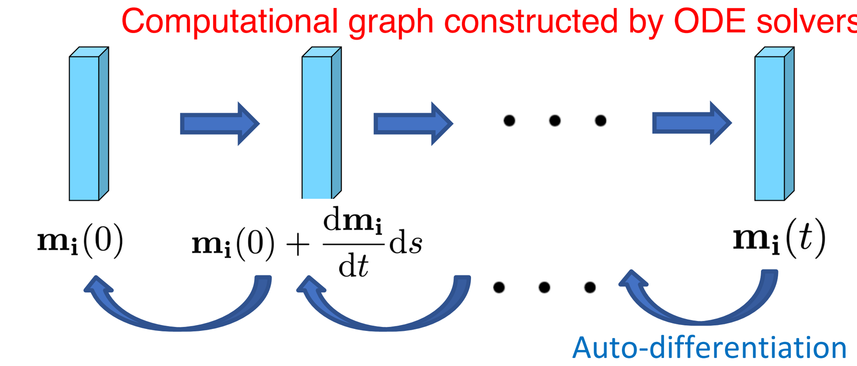
Joint probability given observations $\mathcal{D} = \{(\mathbf{i}_1, t_1, y_1), \ldots, (\mathbf{i}_N, t_N, y_N)\}$

$$p(\mathcal{U}, \nu, \mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \prod_{j=1}^{d_k} \mathcal{N}(\mathbf{u}_j^k | \mathbf{0}, \mathbf{I}) \cdot \text{Gam}(\nu | a_0, b_0) \cdot \prod_{n=1}^{N} \mathcal{N}(y_n | m_{\mathbf{i}_n}(t_n), \nu^{-1})$$

Auto-differentiation or explicit gradients?

Computational graph constructed by ODE solvers

All the intermediate states needs to be saved to compute the gradients



Efficient computation of gradients with *Forward Sensitivity*

$$J(m_{\mathbf{i}_n}(t_n)) = \log p(y_n | m_{\mathbf{i}_n}(t_n)) \xrightarrow{\text{Chain rule}} \frac{dJ}{d\eta} = \frac{\partial J}{\partial \eta} + \frac{\partial J}{\partial m_\mathbf{i}(t)} \cdot \frac{d m_\mathbf{i}(t)}{d\eta}$$

Easy to compute

$$\frac{d m_\mathbf{i}(t)}{d\eta} = \frac{\partial m_\mathbf{i}(t)}{\partial \eta} + \frac{\partial m_\mathbf{i}(t)}{\partial m_\mathbf{i}^0} \cdot \frac{d m_\mathbf{i}^0}{d\eta}$$

Sensitivities of the system

Def aux state $\mathbf{z_i} = [m_\mathbf{i}^0; \boldsymbol{\eta}]$

$$\mathbf{s_i}(t) = \frac{\partial m_\mathbf{i}(t)}{\partial \mathbf{z_i}} = \left[\frac{\partial m_\mathbf{i}(t)}{\partial m_\mathbf{i}^0}; \frac{\partial m_\mathbf{i}(t)}{\partial \boldsymbol{\eta}}\right]$$

Now to derive the dynamics of the sensitivity

Take time derivative again on sensitivity

$$\frac{d}{dt}\underbrace{\frac{\partial m_\mathbf{i}(t)}{\partial \mathbf{z_i}}}_{\dot{\mathbf{s}}_\mathbf{i}(t)} = \frac{\partial}{\partial \mathbf{z_i}} \frac{d m_\mathbf{i}(t)}{dt} = \frac{\partial f}{\partial m_\mathbf{i}(t)} \underbrace{\frac{\partial m_\mathbf{i}(t)}{\partial \mathbf{z_i}}}_{\mathbf{s}_\mathbf{i}(t)} + \frac{\partial f}{\partial \mathbf{z_i}}$$

ODE of system sensitivity

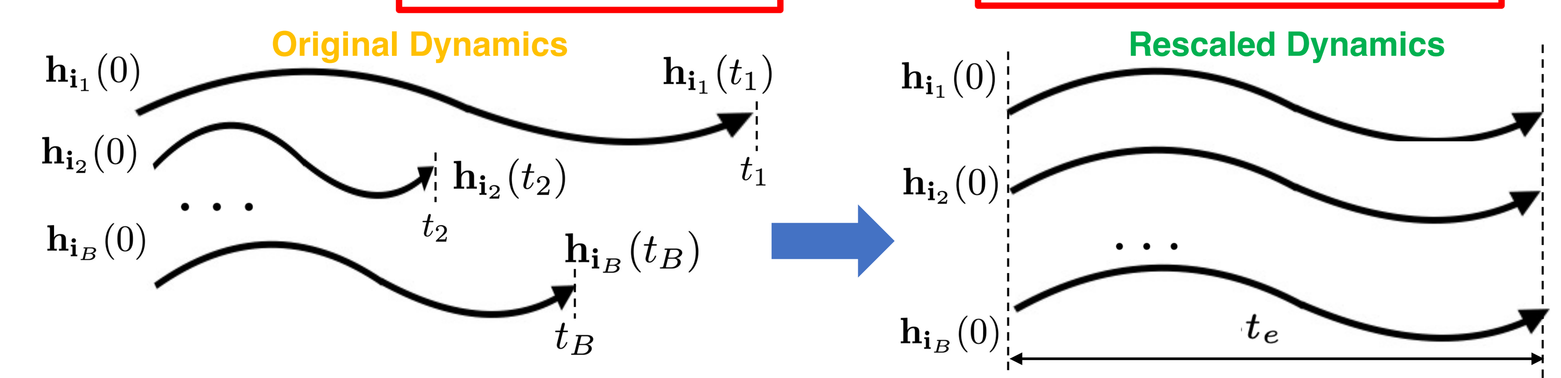
$$\begin{cases} \frac{d\mathbf{s_i}(t)}{dt} = \frac{\partial f}{\partial m_\mathbf{i}(t)} \mathbf{s_i}(t) + \frac{\partial f}{\partial \mathbf{z_i}} \\ \mathbf{s_i}(0) = \frac{d m_\mathbf{i}^0}{d\mathbf{z_i}}, \end{cases}$$

- Depends on current state solution only $\quad \mathbf{h_i}(t) = [m_\mathbf{i}(t); \mathbf{s_i}(t)]$
- **Jointly solved** with the system state with ***and forward pass only one ODE solver***

Time alignment for efficient stochastic mini-batch optimization

$$\widehat{\mathcal{L}} = \log(\text{Prior}) + \frac{N}{B} \sum_{j=1}^{B} J(m_{\mathbf{i}_{n_j}}(t_{n_j}))$$

$$\mathbf{h}_{\mathbf{i}_l}(t_l) = \mathbf{h}_{\mathbf{i}_l}(0) + \int_0^{t_l} \boldsymbol{\alpha}(\mathbf{h}_{\mathbf{i}_l}(\tau), \tau) d\tau = \mathbf{h}_{\mathbf{i}_l}(0) + \int_0^{t_e} \frac{t_l}{t_e} \boldsymbol{\alpha}\left(\mathbf{h}_{\mathbf{i}_l}(\frac{t_l}{t_e}s), \frac{t_l}{t_e}s\right) ds$$



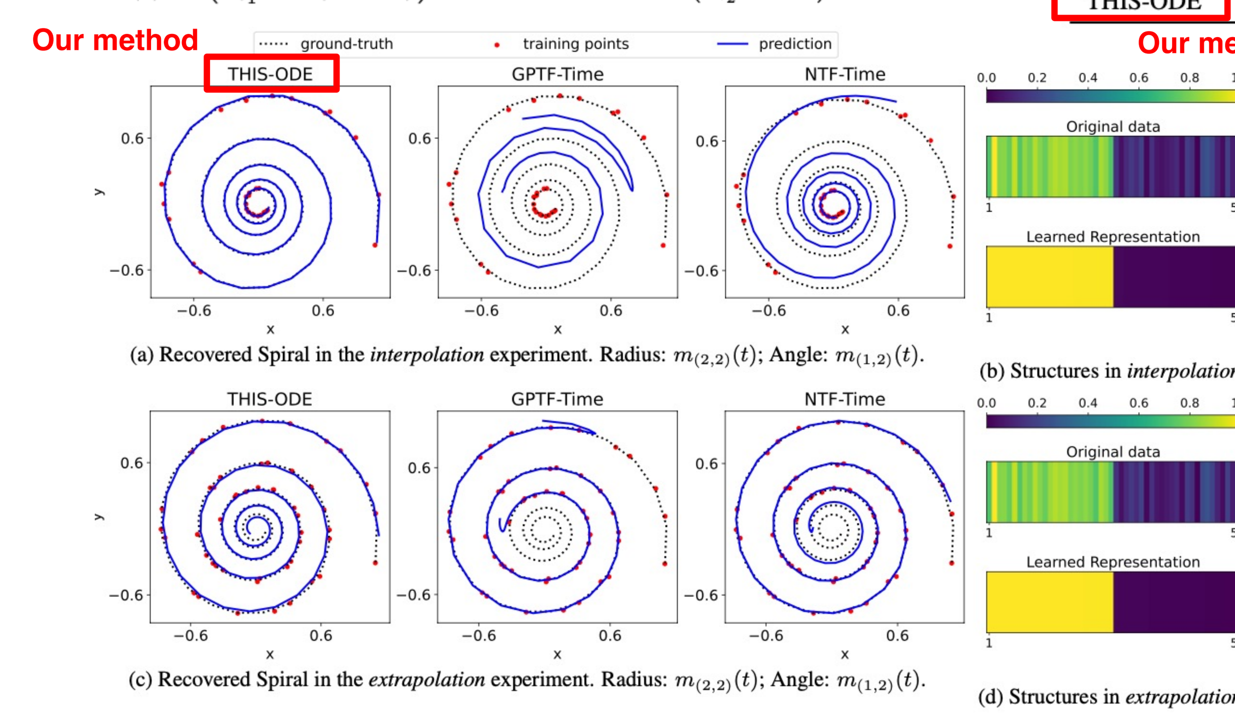
Experiment

Ablation Study: Spiral Interactions

$$m_\mathbf{i}(t) = (u_{i_1}^1 \exp(-0.5t))^{\mathbb{1}(i_1+i_2 \mod 2=0)} \cdot (u_{i_2}^2 + 2\pi t)^{\mathbb{1}(i_1+i_2 \mod 2=1)}$$

	Interpolation	Extrapolation
GPTF-Time	0.5557	0.9032
NTF-Time	0.1004	0.3656
THIS-ODE	**0.0148**	**0.0746**

Our method



(a) Recovered Spiral in the *interpolation* experiment. Radius: $m_{(2,2)}(t)$; Angle: $m_{(1,2)}(t)$.

(b) Structures in *interpolation*.

(c) Recovered Spiral in the *extrapolation* experiment. Radius: $m_{(2,2)}(t)$; Angle: $m_{(1,2)}(t)$.

(d) Structures in *extrapolation*.

Ablation Study: Spiral Interactions: *Beijing Air, Indoor Condition, Fit Record, Server Room …*

Name	Description	Size	NNZ	Granularity in Time
Beijing Air Quality	time x locations x pollutants	35064 x 12 x 6	2454305	hourly
Indoor Condition	time x locations x sensor	19735 x 9 x 2	241201	every 10 minutes

Interpolation	*Beijing Air*	*Indoor Condition*
CP-Time	0.897 ± 0.012	0.780 ± 0.012
CP-DTL	0.898 ± 0.015	0.842 ± 0.003
CP-DTN	0.833 ± 0.003	0.889 ± 0.005
GPTF-Time	0.711 ± 0.011	0.849 ± 0.005
GPTF-DTL	0.686 ± 0.045	0.852 ± 0.004
GPTF-DTN	0.670 ± 0.062	0.713 ± 0.104
NTF-Time	0.745 ± 0.095	0.800 ± 0.009
NTF-DTL	0.757 ± 0.006	0.777 ± 0.018
NTF-DTN	0.686 ± 0.011	0.665 ± 0.004
PTucker	0.959 ± 0.015	0.806 ± 0.027
THIS-ODE	$\mathbf{0.624 \pm 0.008}$	$\mathbf{0.618 \pm 0.007}$

. . .

Extrapolation		
CP-Time	0.863 ± 0.022	0.867 ± 0.010
CP-DTL	0.553 ± 0.005	0.527 ± 0.006
CP-DTN	0.557 ± 0.004	0.584 ± 0.009
GPTF-Time	0.527 ± 0.018	0.489 ± 0.011
GPTF-DTL	0.577 ± 0.035	0.506 ± 0.013
GPTF-DTN	0.511 ± 0.002	0.489 ± 0.003
NTF-Time	0.537 ± 0.002	0.510 ± 0.027
NTF-DTL	0.512 ± 0.009	0.593 ± 0.079
NTF-DTN	0.513 ± 0.003	0.484 ± 0.011
PTucker	0.522 ± 0.022	0.749 ± 0.006
THIS-ODE	$\mathbf{0.498 \pm 0.013}$	$\mathbf{0.460 \pm 0.004}$

. . .

Our method