Nonparametric Embeddings of Sparse High-Order Interaction Events

Zheng Wang[1], Yiming Xu[2], Conor Tillinghast[2], Shibo Li[1], Akil Narayan[2], Shandian Zhe[1]

School of Computing[1], Department of Mathematics[2]

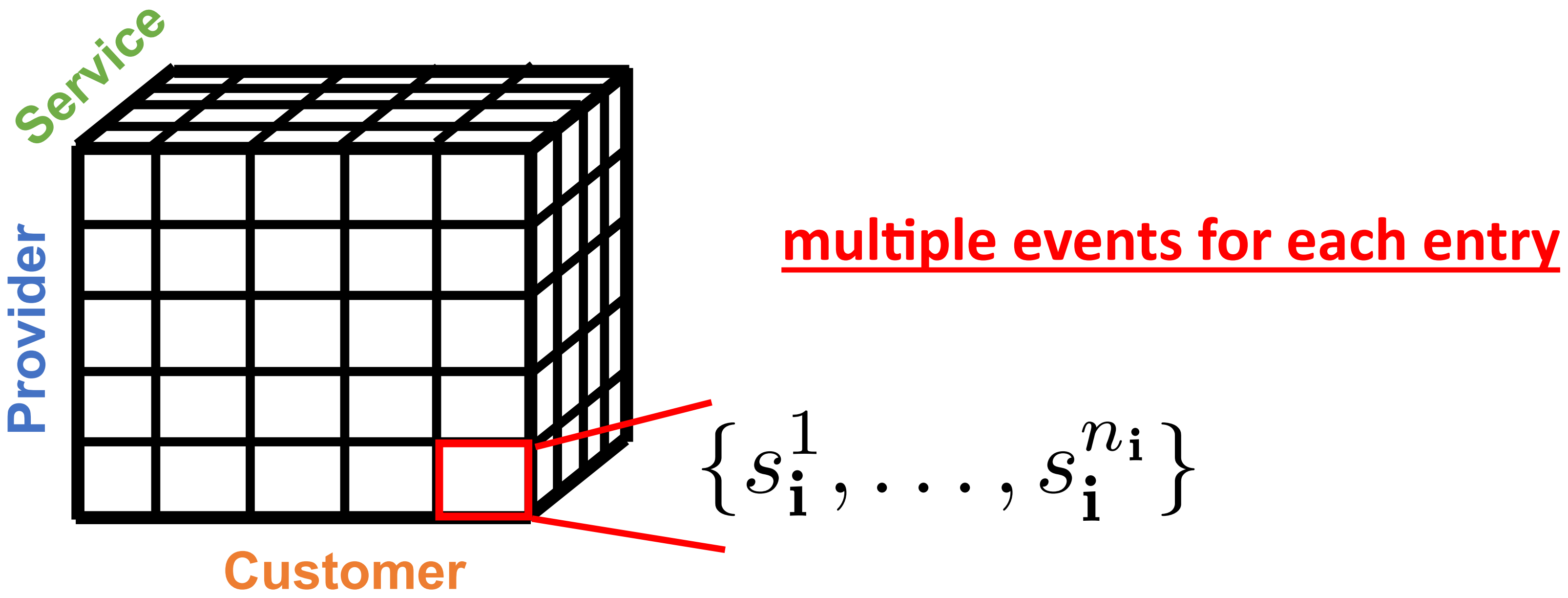
University of Utah

UNIVERSITY OF UTAH

Quick Overview

$$S = [(s_1, \mathbf{i}_1) \ldots, (s_N, \mathbf{i}_N)]$$

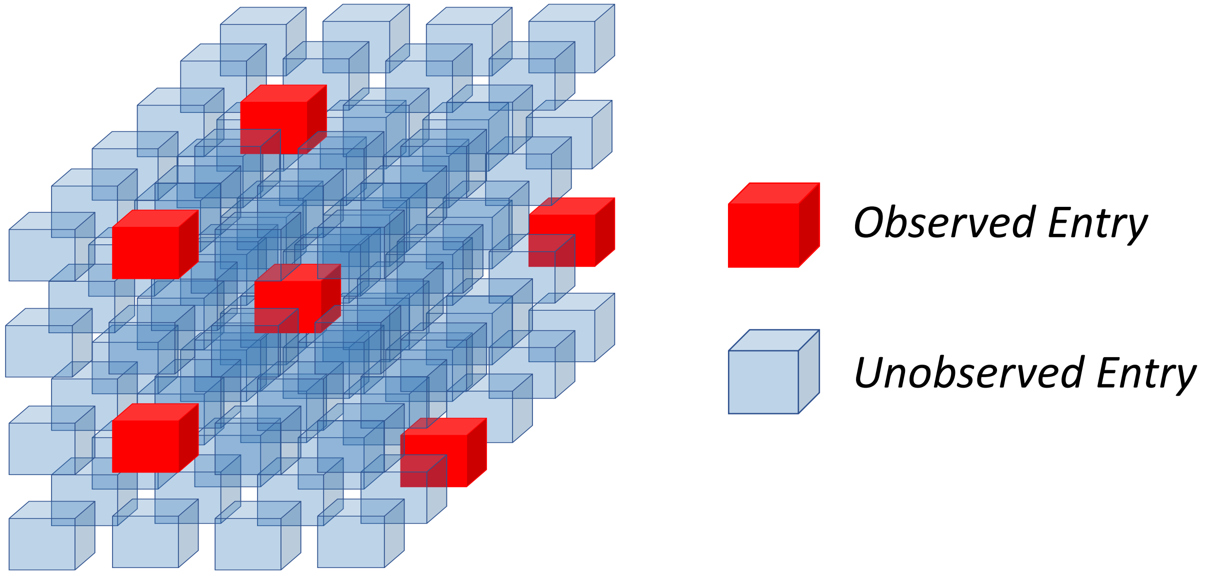
Event Tensor



$$\{s_{\mathbf{i}}^{1}, \ldots, s_{\mathbf{i}}^{n_{\mathbf{i}}}\}$$

- Interactions(events) in real-world applications often involve multiple entities
- *For example, consumers' activities can be viewed as interactions between customers, service items and providers*

Overlooked Sparsity



- The underlying sparse structure of data: the occurred interactions are ***far less than*** the possible interactions among all the participants.
- Current Poisson tensor factorization methods ***ignore the sparsity*** and assume dense tensor structures.
- ***Sparse tensor (hyper- graph) processes (STP)*** (Tillinghast and Zhe, 2021) provides an efficient way to incorporate the sparsity as a prior.

Nonparametric Embeddings of Sparse High-Order Interaction Events

- ***Major Contributions:***
  - **Model:** We hybridize the recent sparse tensor (hyper-graph) processes (STP) (Tillinghast and Zhe, 2021) and matrix Gaussian processes (MGP) to develop a ***sparse event model***.
  - **Theory:** We use Poisson tail estimate, Bernstein's inequality and L'Hôpital's rule to prove ***strong asymptotic bounds of the sparsity ratio***, including both a lower and upper bound.
  - **Algorithm:** We use the stick-breaking construction of the normalized hypergraph process to compute the embedding prior, and then use batch-normalization and variational sparse GP framework to develop an ***efficient and scalable*** model estimation algorithm.

Sparse Structure Prior – STP

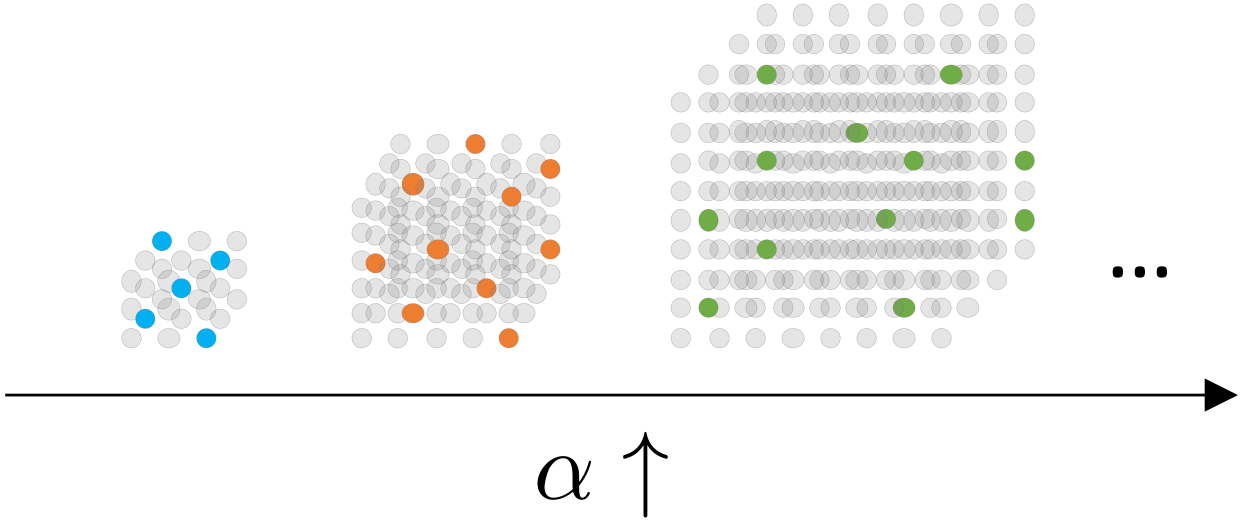
$$G_r^k \sim \mathrm{DP}\big(\alpha, \mathrm{Uniform}([0, \alpha])\big),$$

$$\widehat{G}_r^k = \sum_{j=1}^{\infty} \omega_{rj}^k \cdot \delta_j.$$

$$\widehat{M} = \sum_{\mathbf{i}=(1,\ldots,1)}^{(\infty,\ldots,\infty)} w_{\mathbf{i}} \cdot \delta_{\mathbf{i}},$$

$$p(\mathcal{E}) = \prod\nolimits_{n=1}^{N} p(\mathbf{i}_n) = \prod\nolimits_{n=1}^{N} w_{\mathbf{i}_n}.$$

Lemma 3.1 (Corollary 3.1.1 (Tillinghast and Zhe, 2021)). $N^\alpha = o(\prod_{k=1}^{K} D_k^\alpha)$ *almost surely as* $\alpha \to \infty$, *i.e.*, $\lim_{\alpha\to\infty} \frac{N^\alpha}{\prod_{k=1}^{K} D_k^\alpha} = 0 \quad a.s.$



Matrix Gaussian Process

- Matrix Gaussian process (MPG) (Rasmussen and Williams, 2006) for rate function

$$\boldsymbol{\rho} \sim \mathcal{MN}(\mathbf{0}; \kappa_1(\mathbf{x_i}, \mathbf{x_{i'}}), \kappa_2(t, t')),$$

$$p(\mathcal{S}|\boldsymbol{\lambda}) = \prod_{n=1}^{N} \exp(-\int_0^T (\rho_{\mathbf{i}_n}(t))^2 \,\mathrm{d}t) \prod_{j=1}^{m_{\mathbf{i}_n}} (\rho_{\mathbf{i}}(s_{\mathbf{i}_n j}))^2, \quad (9)$$

Theoretical Analysis of Sparsity

- ***Strong asymptotic bounds*** of the ***sparsity ratio***, including both a lower and upper bound

$$W_{k,r}^{\alpha} \sim \Gamma\mathrm{P}(\beta_{\alpha}) \quad (1 \le k \le K, 1 \le r \le R)$$
$$M = \sum_{r=1}^{R} W_{1,r}^{\alpha} \times \ldots \times W_{K,r}^{\alpha},$$
$$T | \{W_{k,r}^{\alpha}\}_{1 \le k \le K, 1 \le r \le R} \sim \mathrm{PPP}(M). \qquad (3)$$

Lemma 3.2. *For a sparse hyper-graph process defined as in (3), for all sufficiently large α, there exists an absolute constant $C > 0$ such that, with probability at least $1 - (C\alpha)^{-K}$,*

$$\frac{e^{-1.03(2K)^{1/K} K (\log \alpha)^{1/K}}}{2K \log \alpha} \cdot \left[\frac{1.82}{(K-1)\log(1.01\alpha)}\right]^{K} \le \frac{N^{\alpha}}{\prod_{k=1}^{K} D_k^{\alpha}} \le \left[\frac{2.11}{(K-1)\log(0.99\alpha)}\right]^{K}.$$

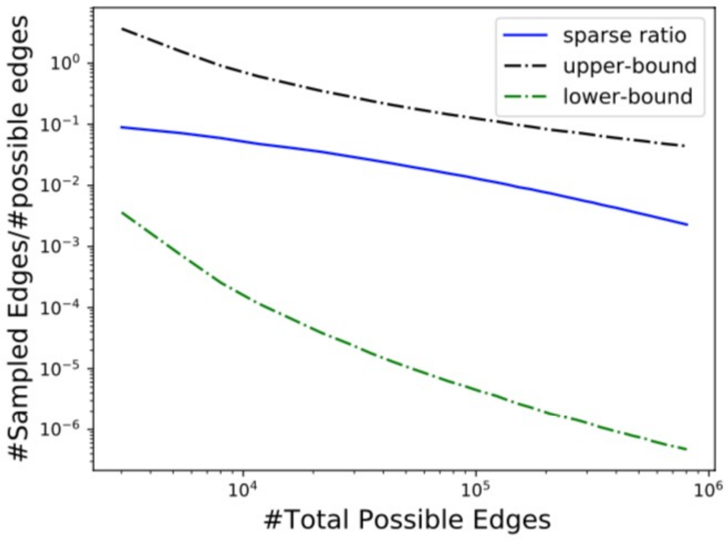
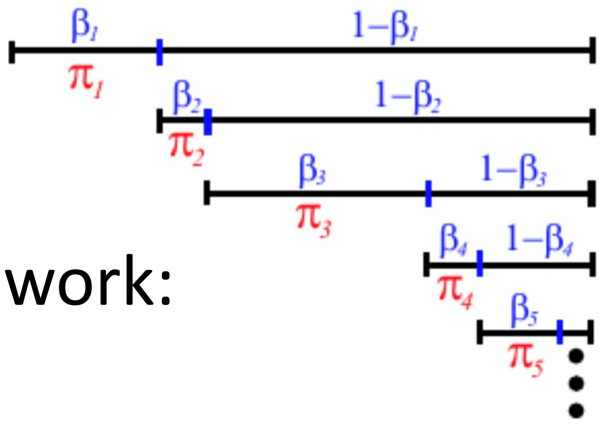


Figure 1: Sparsity ratio of the sampled hypergraphs and bounds.

Algorithm – Model Inference

- Stick-breaking construction for sparse prior

$$v_{rj}^k \sim \text{Beta}(1, \alpha), \quad \omega_{rj}^k = v_{rj}^k \prod_{l=1}^{j-1} (1 - v_{rl}^k),$$

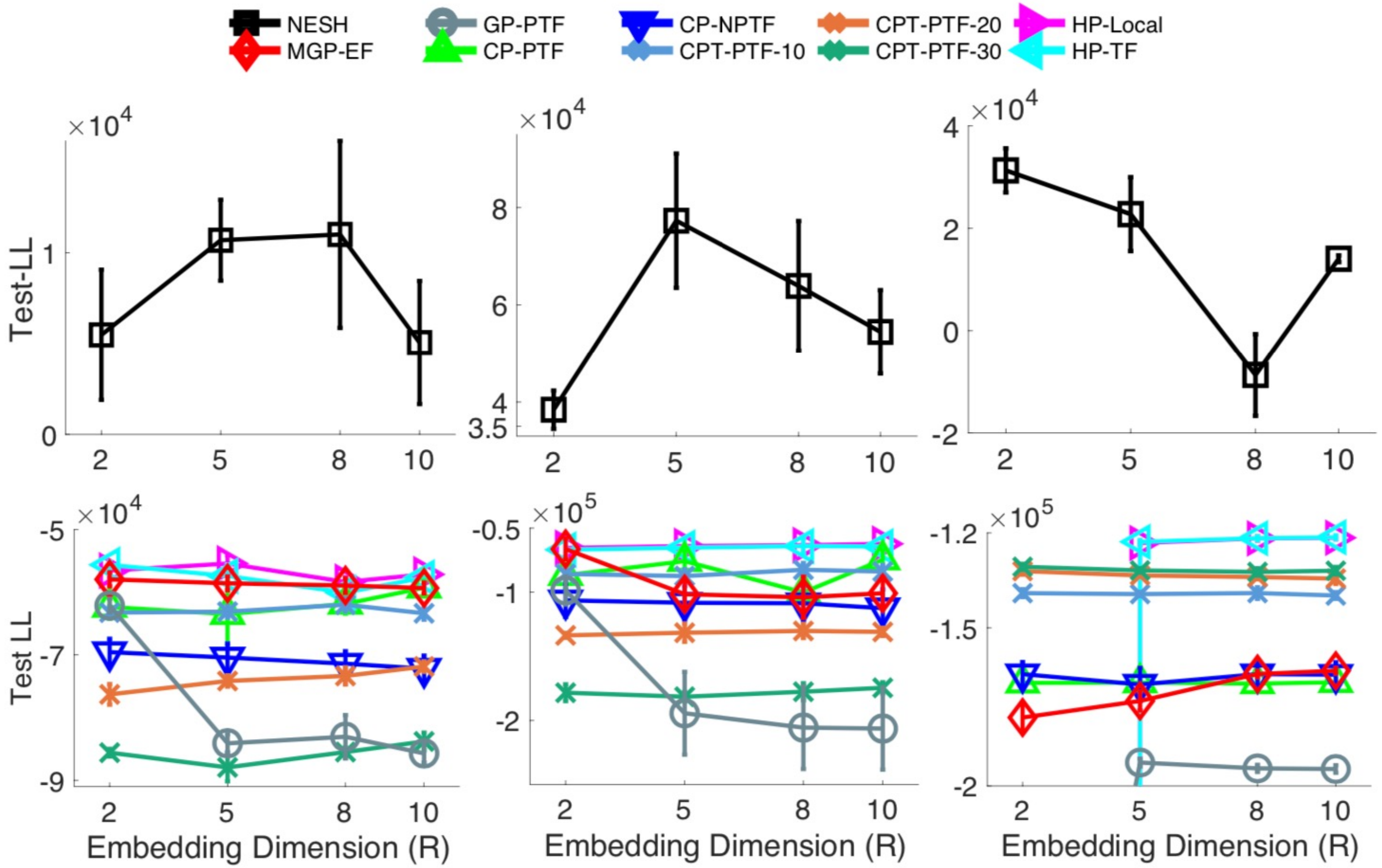


- ELBO with sparse variational Gaussian process framework:

$$\mathcal{L} = \mathbb{E}_{q(\mathbf{b},\mathbf{f})}\left[\log \frac{p(\text{Joint})}{q(\mathbf{b},\mathbf{f})}\right]$$
$$= \mathbb{E}_q\left[\log \frac{\text{OtherTerms} \cdot p(\mathbf{b})\cancel{p(\mathbf{f}|\mathbf{b})}p(\mathcal{S}|\mathbf{f})}{q(\mathbf{b})\cancel{p(\mathbf{f}|\mathbf{b})}}\right].$$

- Stochastic mini-batch optimization

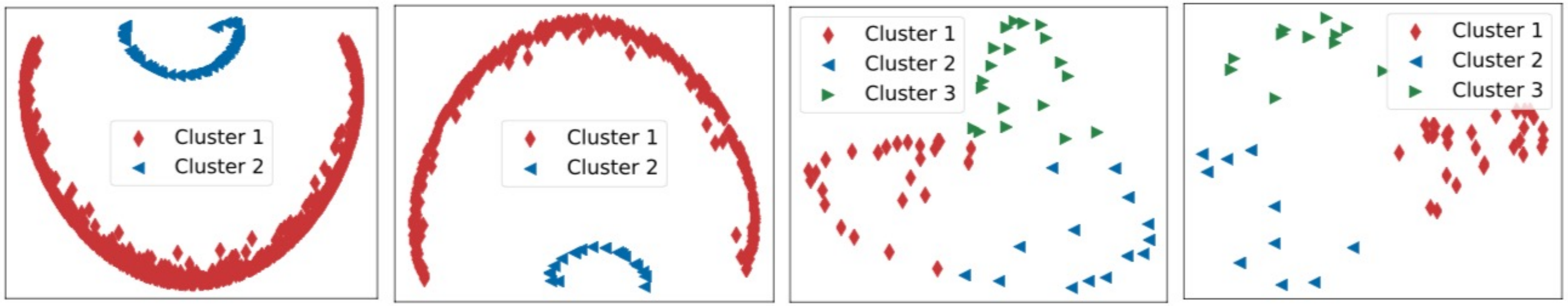
Experiments – Loglikelihood



(a) *Crash* (b) *Taobao* (c) *Retail*

Figure 2: Test log-likelihood (LL) on real-world datasets. CPT-PTF-{10, 20, 30} means running CPT-PTF with 10, 20 and 30 time steps. The results were averaged over five runs.

Experiments – Pattern Discovery



(a) *Users* (b) *Sellers* (c) *Item Categories* (d) *States*

Figure 3: Structures of the estimated embeddings on *Taobao* (a, b, c), *Crash* (d). The points represent the participant nodes, and the colors indicate their cluster memberships.

Conclusion

- We have presented NESH, a ***novel nonparametric embedding method*** for sparse high-order interaction events. Not only can our method estimate the complex temporal relationships between the participants, our model is also able to capture the structural information underlying the observed ***sparse interactions***. Our theoretical bounds enable ***convergence rate estimate*** and reveal insights about the ***asymptotic behaviors of the sparse prior*** over hypergraphs or tensors.

Thanks!

wzhut@cs.utah.edu
University of Utah