Discovering Latent Covariance Structures for Multiple Time Series

Anh Tong and Jaesik Choi

Ulsan National Institute of Science and Technology

Introduction

- **Goal:** extract explainable representations (temporal covariance) shared among multiple inputs (time series)
- **Our contributions**:
    - **Latent Kernel Model (LKM)**: a new combination of two nonparametric Bayesian methods handling multiple time series
    - **Partial Expansion (PE)**: an efficient kernel search for multiple inputs
    - Automated reports emphasizing the characteristics of individual data

Two nonparametric methods

- Gaussian process (GP): prior over function values

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

  Important to choose an **appropriate** kernel

- Indian Buffet Process (IBP): prior over binary matrices

Finite (Beta-Bernoulli) $\Longrightarrow$ Infinite (IBP)

$$z_{nk} \sim \text{Bernoulli}(\theta_k)$$

$$\theta_k \sim \text{Beta}(\alpha/K, 1)$$

$$\lim_{K\to\infty} P(\mathbf{Z}|\alpha) = 0$$

$$[\mathbf{Z}] = lof(\mathbf{Z})$$

$$P(\mathbf{Z}|\alpha) = \prod_k \frac{\Gamma(m_k+\frac{\alpha}{K})\Gamma(N-m_k+1)}{\Gamma(\frac{\alpha}{K})} \frac{\Gamma(1+\frac{\alpha}{K})}{\Gamma(N+1+\frac{\alpha}{K})}$$

$$\lim_{K\to\infty} P([\mathbf{Z}]|\alpha) = \exp\{-\alpha H_N\} \frac{\alpha^{K_+}}{\prod_{h>0} K_h!} \prod_{k\le K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

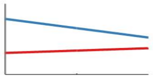
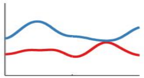
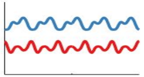
Exchangeability among columns

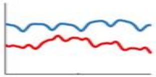
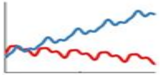
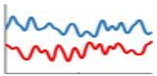
Compositional kernel learning in Automatic Statistician

[Duvenaud et al. 2013]

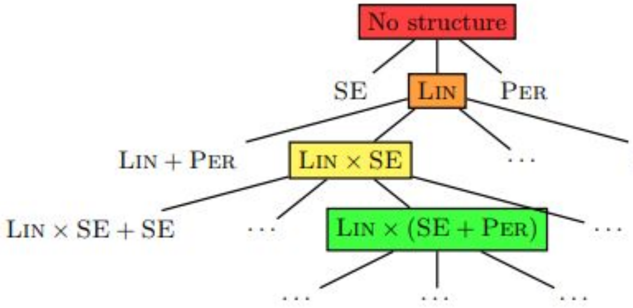
Two main components:

- **Language of models:**
  - Base kernels: SE, LIN, PER
  - Operators: +, x, change point & window

Base kernels	
Linear (LIN)	
Smooth (SE)	
Periodic (PER)	

Kernel composition	
SE+PER	
LIN+PER	
SExPER	

- **Search procedure:**
  - A greedy manner
  - Model is selected based on trade-off between model and data complexity



Relational kernel learning [Hwang et al. 2016] introduced a kernel learning for multiple time series by assuming a globally shared a kernel and individual spectral mixture kernels.

Latent Kernel Model [This paper]

- Construct GP kernels by a sum of kernels with indicator matrix Z

n: index of time series
k: index of explainable kernel

(1) sample from IBP $\mathbf{Z} \sim \mathrm{IBP}(\alpha)$

membership

(2) kernel construction $c_n(t, t') = \sum_k z_{nk} c_k(t, t')$

(3) function values are modeled by GP $f_n(t) \sim \mathcal{GP}(0, c_n(t, t'))$

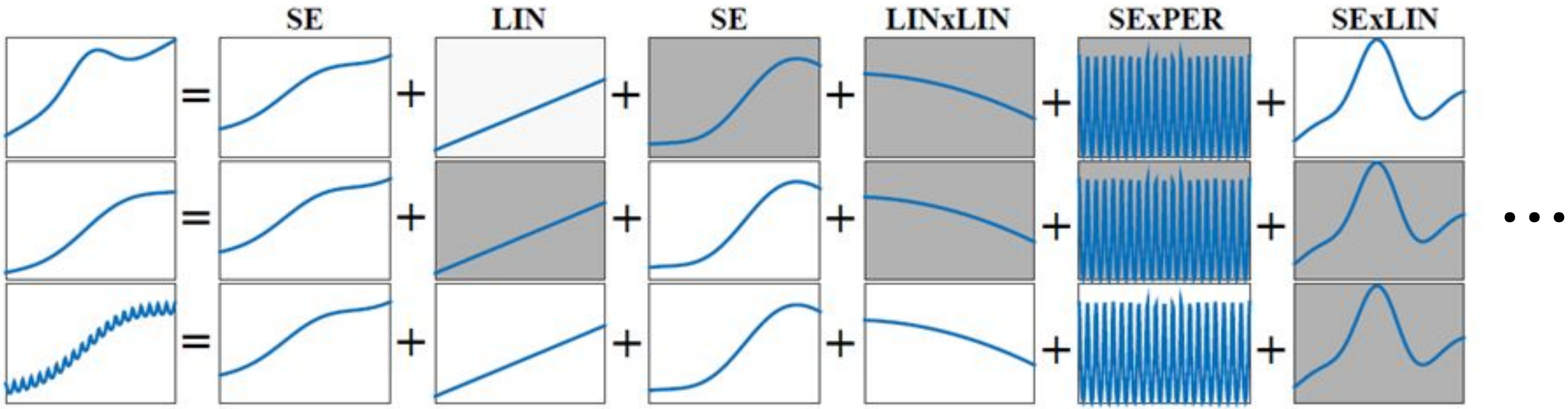
Proposition 1. *With* $\mathbf{Z} \sim \mathrm{IBP}(\alpha)$ *, the likelihood of LKM*

$$p(\mathbf{X}|\mathbf{Z}) = \prod_n \mathcal{N}(\mathbf{x}_n; \mathbf{0}, \mathbf{D}(\mathbf{z}_n))$$

where $\mathbf{D}(\mathbf{z}_n) = \sum_{k=1}^{\infty} \mathbf{C}_k + \sigma_n^2 \mathbf{I}$ *, is well-defined.*

Proof. *We showed with the commutative among additive kernels and the exchangeability of columns (lof).*

Latent Kernel Model [This paper]

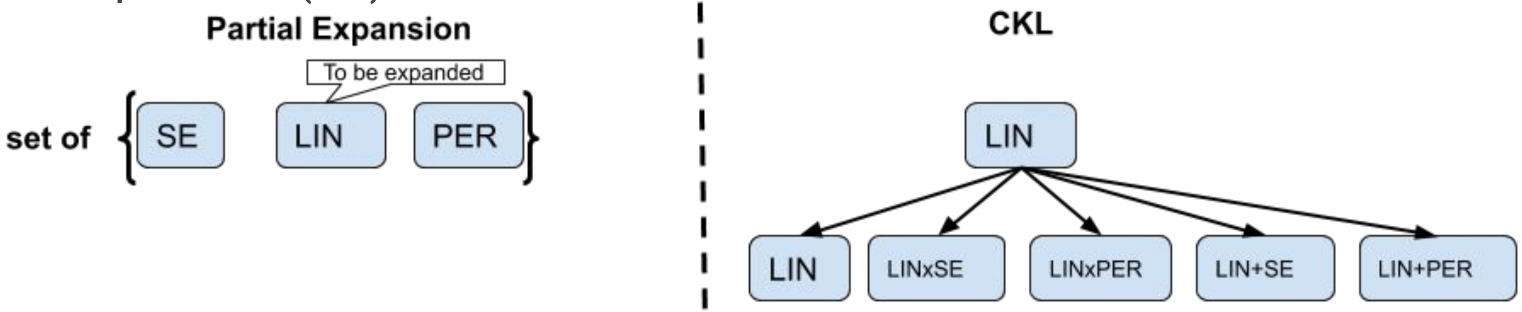


Enlarged covariance structure search

- **Challenge**: CKL cannot directly apply to multiple time series, e.g., a different structure for a time series
- Partial expansion (PE):

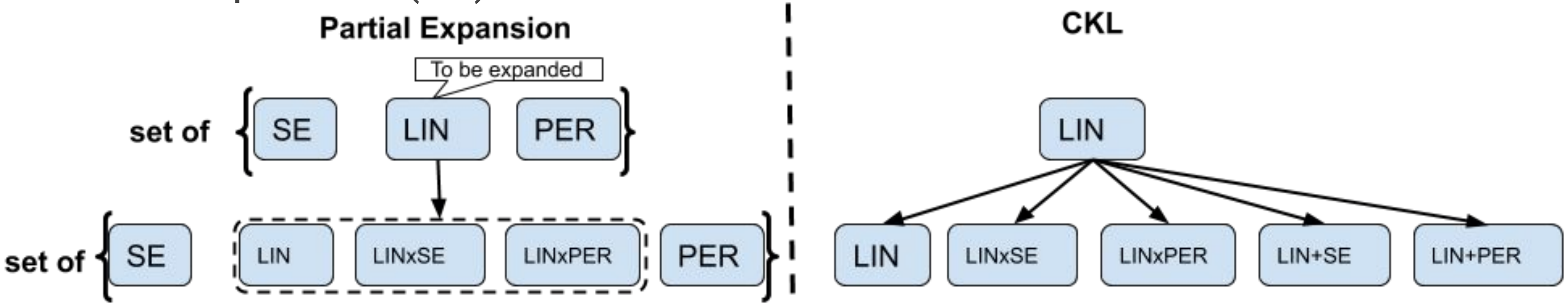
Enlarged covariance structure search

- **Challenge**: CKL cannot directly apply to multiple time series, e.g., a different structure for a time series
- Partial expansion (PE):



Enlarged covariance structure search

- **Challenge**: CKL cannot directly apply to multiple time series, e.g., a different structure for a time series
- Partial expansion (PE):



- Maintain a set of kernels
- Iteratively expand a kernel in the set to obtain a new model
- **Note**: PE explores a larger structure space

Approximate inference

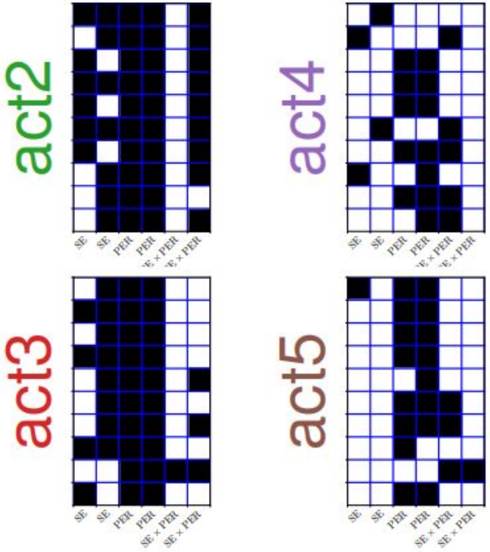
- Maximize the evidence lower bound

$$\log p(\mathbf{X}) \geq \mathbb{E}[\log p(\mathbf{Z})] + \mathbb{E}[\log p(\mathbf{X}|\mathbf{Z})] + H[q]$$

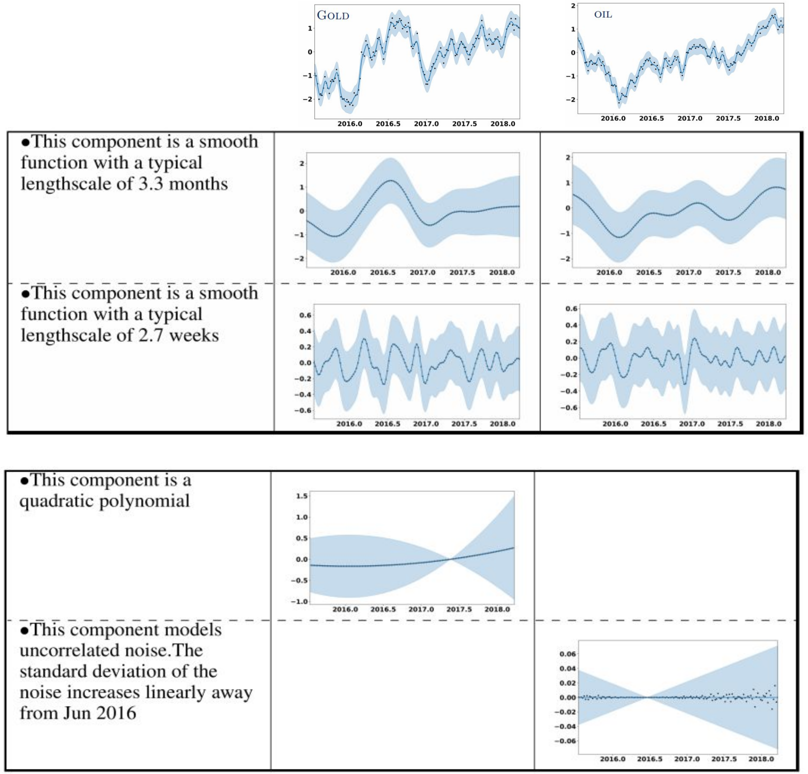
- **Challenge**: Estimating $\mathbb{E}[\log p(\mathbf{X}|\mathbf{Z})]$ is expensive, e.g., # computing Gaussian log-likelihood grows exponentially as **K** increases.

- **Solution:**
  - Relax discrete R.V. to continuous R.V. by reparameterization with Gumbel-Softmax trick
  - Approximate by MCMC

Qualitative demonstration



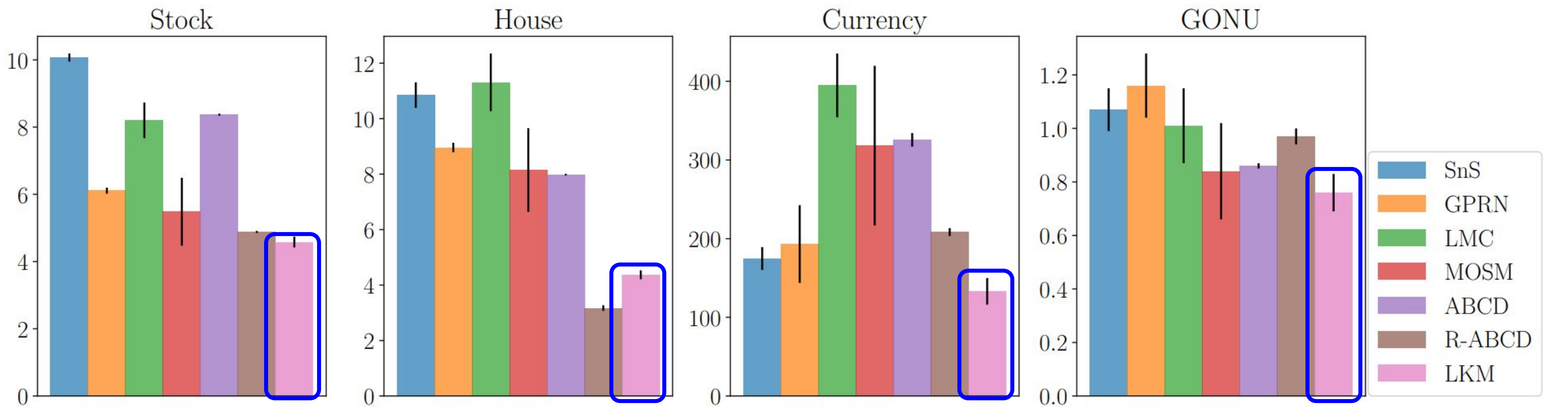
SEIZURE DATA



FINANCIAL DATA

- Interpretability of IBP matrix: reveal characteristics of different activities
- A new type of automatic generated reports taken into account the comparative relations

Quantitative result



- Tested on various data sets, e.g. closely correlated to loosely correlated
- Outperform multi-output and CKL-based methods

Conclusion

- Present a model analyzing and **explaining multiple time series**
- **Improve kernel search procedure** to facilitate model discovery
- Provide **a detailed comparison report**

poster #226